Enlightened Update: A Computational Architecture for Presupposition and Other Pragmatic Phenomena

VERSION 1.0 — JUNE 3, 2006

Richmond H. Thomason[1], Matthew Stone[2,3] and David DeVault[2]
[1]Philosophy [2]Computer Science [3]HCRC
Michigan Rutgers Edinburgh

FOR A VOLUME EDITED BY BYRON, ROBERTS AND SCHWENTER

Abstract

We relate the theory of presupposition accommodation to a computational framework for reasoning in conversation. We understand presuppositions as private commitments the speaker makes in using an utterance but expects the listener to recognize based on mutual information. On this understanding, the conversation can move forward not just through the positive effects of interlocutors' utterances but also from the retrospective insight interlocutors gain about one anothers' mental states from observing what they do. Our title, ENLIGHTENED UPDATE, highlights such cases. Our approach fleshes out two key principles: that interpretation is a form of intention recognition; and that intentions are complex informational structures, which specify commitments to conditions and to outcomes as well as to actions. We present a formalization and implementation of these principles for a simple conversational agent, and draw on this case study to argue that pragmatic reasoning is holistic in character, continuous with common-sense reasoning about collaborative activities, and most effectively characterized by associating specific, reliable interpretive constraints directly with grammatical forms. In showing how to make such claims precise and to develop theories that respect them, we illustrate the general place of computation in the cognitive science of language.

1 Introduction

Computational approaches to pragmatics are informed by pragmatic theories, and shape these theories by demanding explicit accounts of pragmatic reasoning and by providing a laboratory in which these theories can be tested. In this paper, we illustrate this interaction by showing how computationally inspired ideas can help in developing an adequate theory of accommodation.

Our ultimate goal is to develop an explicit, computational model of the reasoning that allows interlocutors to understand each other correctly in specific contexts. Such a model will specify the algorithms that accomplish this reasoning and the symbolic data structures over which the algorithms are defined. But this model will inevitably go beyond the computations themselves. It will, in addition, offer insight into the underlying PROBLEMS that interpretation algorithms need to solve, and into the real-world INFORMATION that pragmatic structures need to represent. As has been long recognized in Cognitive Science (Newell, 1981; Marr, 1982), it is at this "knowledge level", or level of "computational theory", that computational approaches can most directly engage with other methodologies for the study of mind and behavior. Accordingly, we begin this paper with a knowledge-level outline of the approach we will develop. We characterize conversation as a collaboration and characterize pragmatic reasoning in particular as an outgrowth of

speakers' and listeners' collaborative efforts to coordinate on communicative intentions.

1.1 The requirements of conversational coordination

Like many others across Cognitive Science (see, for instance, (Lewis, 1969; Clark, 1992)), we characterize communication as a problem of coordination. In conversation, speakers produce each utterance with a particular interpretation in mind; listeners infer an interpretation for each utterance. Successful communication requires that the interlocutors' interpretations of each utterance agree. At the knowledge level, the problem of coordination places strong constraints on the information interlocutors should use in conversation; they should only use information that is public, or MUTUALLY SUPPOSED (Stalnaker, 1975; Thomason, 1990) in a sense we will describe more fully below. We summarize this by the coordination principle in (1).

(1) In conversations, speakers must formulate, and listeners must infer, an interpretation for each utterance that is uniquely recognizable from the form of the utterance, using mutually-supposed background information.

1.2 Intentions

Meanwhile, again like many others across Cognitive Science inspired by (Grice, 1957), we appeal to the theory of intention to flesh out the content of utterance interpretations. Interactions between Artificial Intelligence and Philosophy have been particularly fruitful in the area of rational agency. These ideas can be applied to conversation, viewed as the rational exchange of information. As in theories of agency along the lines of (Bratman, 1987; Pollack, 1990; Pollack, 1992), we understand an intention as a complex mental attitude that summarizes an agent's reason to act.

Intentions perform a range of functions in rational agency. Intentions can help an agent to manage problem solving in planning, to cope with the obstacles and opportunities that arise in a dynamic world, and to understand why actions fail and how to respond when they do. Of course, intentions can also help agents to work more effectively together. It is because intentions play these varied roles in agents' deliberations that researchers have analyzed intentions as complex informational structures. Intentions delineate the real-world circumstances that the agent is committed to, characterize the action that the agent undertakes, and explain, in the context of the agent's theory of the world, why taking that action in those circumstances will lead to an outcome that the agent wants.

Ideas that again trace back to Grice treat the interpretation process in conversation as a species of INTENTION RECOGNITION—a hearer's reasoning task is to reconstruct the speaker's communicative intention. This, together with the above characterization of intention, yields the following characterization of the interpretations that interlocutors infer in pragmatic reasoning.

(2) Interpretations outline a speaker's commitment not just to the content of a specific contribution to conversation, but also to the underlying intention, and through this to the structure of the utterance (as analyzed by the grammar) and to the relevant extralinguistic facts that link grammatical meaning to conversational content.

Of particular note for our purposes is the fact that utterances, like other actions, may depend for their effects on preconditions about the situation in which they are performed. Thus, as made explicit in (2), a speaker's communicative intention will register apposite commitments about the conversational situation.

Models of intention thus suggest a knowledge-level account of conversation framed in terms of the common-sense information about utterances that is required in a communicative intention. Our formalism provides data structures—complex, structured pragmatic representations—that package up this information; our model of pragmatic reasoning uses these representations to carry out the inferences involved in attributing a communicative intention to a speaker and calculating the import of recognized intentions for interlocutors' ongoing collaboration (Stone, 2004a; Stone, 2004b). Thus, in contrast to many approaches to formal pragmatics, we do not view our formalism as describing the INPUT to higher-level processes of reasoning about intentions and collaboration. It spells out a SUBSTRATE for this pragmatic reasoning and must itself be narrated in high-level common-sense terms.

1.3 Conversationally manifested private commitments

Our view, ultimately, is just that reasoning processes in conversation are continuous with reasoning processes in other kinds of coordinated, collaborative activity. This idea is familiar (Clark, 1992)—but characterizing these processes in computational terms offers new kinds of precision, clarity, and insight into pragmatic phenomena that seem to depend on interlocutors' inferential abilities. We focus in this paper on the relationship among grammatical knowledge, world knowledge and mutual attitudes in presupposition.

Principles (1) and (2) give a special place in pragmatic theory to PUBLICLY RECOGNIZABLE PRIVATE COMMITMENTS. In using an utterance, as articulated in Principle (2), a speaker commits FOR HERSELF that the conversational situation is as the grammar requires for her utterance to contribute the content she wants. As we shall see, the conversational situation may not encompass everything the speaker believes, but neither must it be limited to the interlocutors' public information. What matters in the first instance is just that the speaker herself believes that the conversational situation is how it needs to be, and and commits to this belief. However, Principle (1) also applies. The speaker's commitment is itself PART of the interpretation, and the listener must recognize the interpretation. So according to Principle (1), the speaker must also make sure that this commitment can be recognized from the utterance and other public information.

We can highlight the significance of this somewhat roundabout characterization by observing that most utterances will require the speaker to commit in this way to MANY specific facts. Each fact might of course be public. But even commitment to information that is NOT public can easily be made recognizable. The speaker need only commit ELSEWHERE to enough public information to clue the listener in to the overall picture she has in mind. Principle (1) says only that the speaker must allow the hearer to infer one interpretation as the best, on the basis of mutual background information.

Interpretations are complex structures, and can express commitment to propositions in many ways. Consider Stalnaker's example (3) (Stalnaker, 1975, p. 202):

(3) I have to pick up my cat at the veterinarian.

The best interpretation of (3) may be an intention to communicate the proposition that I need to take my cat to the veterinarian, an intention that contains a commitment to the proposition that I have a cat. This second commitment may well be private at the time of the utterance but recognized transparently from the utterance itself. Here the speaker expresses an intention to communicate one (private) proposition, and at the same time reveals, as part of this intention, a background commitment to another (private) proposition.

1.4 Abductive reasoning and enlightened update in conversation

This knowledge level account follows naturally from the central idea that interpretation is a form of intention recognition, and that intentions are complex informational structures. But since it is meant to be a computationally useful account, it requires and must inspire an explicit, implementable model of the associated reasoning. Here, we rely especially on models of conversational reasoning developed by Jerry Hobbs and colleagues at SRI—models of interpretation as abduction (Hobbs, Stickel, Appelt & Martin, 1993). Hobbs treats interpretation as a form of inference to the best explanation. His colleague Mark Stickel showed how this reasoning could be implemented in a logic programming framework, and Hobbs showed how the implementation could be used to model a broad sample of pragmatic phenomena.

According to the traditional characterization, abductive reasoning leads from evidence to a hypothesis that explains the observation. An abductive reasoner with a stock of beliefs will observe a proposition E, and add this proposition to its stock of beliefs along with a proposition H, where H is the best explanation of E. We can model interpretation as abduction by taking E to be the observation that an utterance has been made and taking H to specify the speaker's communicative intention in using that utterance. In particular, in keeping with Principle (2), we understand H to spell out the commitments the speaker has made about the initial conversational situation, the grammatical analysis of the utterance, and the consequent contribution that the utterance is to make (Stone & Thomason, 2002).

The revealed intention is thus a rational reconstruction containing much detail that the speaker may not have consciously attended to. Rational reconstructions of this sort are another idea that, in pragmatics at least, go back to Grice. But they are also common in applications of cognitive architectures to nonconversational reasoning.

Abduction can be implemented from a probabilistic standpoint, as a form of statistical inference; but it can also be implemented in more qualitative way, treating explanation as a form of theorem proving that allows least-cost hypotheses to be added as axioms. It is very natural to combine this idea with logic programming methods, which use rules (logical conditionals conforming to certain syntactic restrictions) and data (atomic formulas) to search backwards from a goal to a proof of the goal from the data using the rules.

An abductive theory consists of rules and data, but the rules associate COSTS with their premises. The theorem-proving algorithm of a logic programming method like Prolog is then modified to search through ABDUCTIVE PROOFS— proofs that may involve additional hypotheses—to find the least-cost abductive proof of the goal. The cost of a

proof will depend not only on the costs associated with the rules, but on the available data, and the scoring method favors UNIFYING EXPLANATIONS—hypotheses that explain several different observations at once.

Mark Stickel (Stickel, 1991) developed an algorithm of this sort, which was used as a reasoning engine by Hobbs and his associates. We have modified Stickel's algorithm to allow speech acts to manipulate the weights, which provides a mechanism for managing attention in discourse; see Stone & Thomason (2002). But abductive theorem proving can also be used to form plans—the actions involved in a plan are the "hypotheses" that are invoked to "explain" the goal of a plan. This makes it possible to use abduction as a unified architecture for both interpretation and generation in discourse, and to show explicitly how the architecture enables interlocutors to coordinate on interpretations as required by Principle (1). See Stone & Thomason (Stone & Thomason, 2003).

Abduction is inference to the best OVERALL explanation, taking all the data into account. This methodology requires a holistic approach to pragmatic reasoning. Disparate pragmatic phenomena, such as disambiguation, anaphora resolution, and implicature detection, must be considered together to produce an overall preference for a single, best interpretation in the reasoning process that leads from a contextualized utterance to an update of interlocutors' information.

Although we envisage a unified model of pragmatic reasoning that applies to the listener as well as to the speaker, and so can explain how the reasoning is coordinated in successful communication, our suggestions can best be recapitulated by concentrating on the listener's process of interpretation.

The listener entertains a set of candidate interpretations in attempting to understand an utterance. Each such interpretation is an intention, and so characterizes the speaker as committed to certain (private) claims about the conversational situation, and to a certain desired contribution to the conversation. The listener will draw only on public information in working out which candidate interpretation the speaker must have had in mind. But, if coordination is successful, the outcome will be one preferred interpretation that captures the reason for the speaker's utterance. Significantly, this interpretation may highlight speaker beliefs about the conversational situation of which the listener was previously unaware. And it will usually be unproblematic for the listener to accept these background beliefs, just as it will usually be unproblematic for the listener to accept the speaker's main contribution. Our title, ENLIGHTENED UPDATE, highlights such cases, where the conversation moves forward not just through the positive effects of interlocutors' utterances but also from the retrospective insight interlocutors gain about one anothers' mental states from observing what they do.

1.5 Plan of the paper

Our notion of a publicly recognizable private commitment offers a new way to think about the content of pragmatic judgments. We will argue that this notion brings with it a number of advantages over alternative, knowledge-level accounts of presupposition, because it relates presupposition so directly to the requirements of collaboration and coordination in conversation.

- It allows us to square intuitions about grammatical meaning with intuitions about pragmatic attitudes. We can see, for example, how the grammar might require a speaker to commit to certain information, privately, but in a publicly recognizably way, WITHOUT thereby requiring the speaker to somehow treat it as public, prior information. This gives an attractive way to resolve the well-known and frequent gaps where information must be grammatically backgrounded but need not be shared information among interlocutors. Classic examples include the "informative presuppositions" of change-of-state verbs, factives, and definite noun phrases (Hobbs et al., 1993; Simons, 2001; Roberts, 2003; Abusch, 2005).

- It allows us to narrate pragmatic inference consistently in intuitively appealing ways. For example, we can carry over much of the elegant formal treatment of context change in the presence of accommodation, as developed by authors such as Beaver (Beaver, 2001) and Barker (Barker, 2002), which is couched in terms of uncertainty about the context.[1]

  We can now link this account to models of collaboration and the associated reasoning about real-world task domains in ways that explain HOW a listener can come to be uncertain about what a speaker might recognizably commit to in an utterance. This places the account on a firmer footing by connecting the formalism to an independently-motivated knowledge-level characterization of the problem of interpretation, and clarifying when interlocutors will understand accommodation as an unmarked strategy and when accommodation will trigger marked interpretive effects such as implicatures or repairs.

- Finally, it allows us to IMPLEMENT straightforward models of pragmatic reasoning that respond to a wide range of theoretical and empirical considerations. In fact, we were led to the account in part as a natural outgrowth of our ongoing efforts to implement a collaborative approach to language use (Stone & Thomason, 2003; DeVault et al., 2005). We support our discussions with a description of pragmatic reasoning in a simple implemented conversational agent. We have found that by modeling uncertainty in speakers' possible commitments, while making sure that their actual commitments can be resolved by public information, we have been able to carry over important insights from previous work on collaborative dialogue (Larsson & Traum, 2000; Stone et al., 2001; Purver, 2004) while achieving more flexible and more natural interaction.

[1] There remains, however, room for many positions on how much uncertainty there is about the context in the course of a conversation, and about how mutual attitudes can be supported in the presence of this sort of uncertainty. Our own position can be regarded as a special case of Beaver's very general framework—a special case that places the following three constraints on allowable uncertainty. (1) In our model, the preferences that rank alternative contexts are shared by the conversants and are systematically related to the (abductive) reasoning that supports generation and interpretation. (2) Hence, we can assume that normally, the speaker will minimize ambiguity by presenting the hearer with a most favored alternative interpretation, and that when the speaker fails to do this a clarification subdialogue is likely to ensue. (3) This enables the further assumption that the dialogue returns regularly to CHECKPOINTS, where mutuality is achieved or at least assumed.

The remainder of the paper presents our theory of enlightened update in more detail. Section 2 situates our project, by introducing some received analyses of pragmatic phenomena from Linguistics and the Philosophy of Language, and showing how holistic models of interpretation from Artificial Intelligence offer a new way to reconcile the key intuitions at play. Section 3 motivates the particular approach we take through a fine-grained analogy between enlightened update in silent collaboration on physical tasks and enlightened update in cooperative conversation. Section 4 makes this approach concrete: we sketch the architecture, algorithms, and implemented behavior of a conversational agent that uses enlightened update to interactively identify visual objects with a human user, in a "collaborative reference" task similar to that explored by Clark and colleagues (Clark & Wilkes-Gibbs, 1986; Brennan & Clark, 1996). Finally, Section 5 builds on the perspective we articulate in Sections 3 and 4 to show how the new distinctions we draw can inform and clarify ongoing debates in the theory of presupposition.

We are excited about the rapprochement we envision between theoretical analyses and AI models. Hopefully, this combination of forces can not only bring more coherent theories but can leverage theoretical insights more directly in the design and construction of conversational systems.

2 Background

The literature distinguishes two explanatory roles for pragmatic theory, within which we must frame any theory of presupposition.

The first challenge is to characterize the interface between semantics and pragmatics. It is a truism that our knowledge of linguistic meaning underdetermines the content that an utterance will convey on a particular occasion. Pragmatics is what gets us from meaning to content. A good way to conceptualize the connection is that pragmatic reasoning will start from an underspecified meaning containing free parameters and will derive fully-specified content by instantiating parameters to suitable values retrieved from the conversational situation. The values cannot be retrieved arbitrarily, because they are in fact generally subject to grammatically-specified constraints. Thus, the theory of presupposition must have a place for such constraints, and we survey what we know of such an account in Section 2.1.

The second challenge is to fit pragmatic reasoning into a broader account of communication. In particular, successful communication requires interlocutors to AGREE on the content that is contributed to conversation. Insofar as recognizing the content requires interlocutors to draw on additional background information, the information they actually use must be public. Otherwise, interlocutors cannot be sure they are using the information their partner will expect; miscoordination becomes a possibility. Thus, the theory of presupposition must have a place for the public information that plays this distinctive role in coordination, and we survey the principles of such an account in Section 2.2.

In general, we accept the intuitions and frameworks through which these two challenges have been met in the literature. Except, as we argue in Section 2.3, we reject the common assumption that the same mechanisms are at play in the two cases; we distinguish between the grammatically-specified information that LINKS content to meaning in the conversational situation and the public information that interlocutors use to IDEN-

TIFY content from meaning in a coordinated way. A major constraint on our theory, we argue in Section 2.4 (and on any theory that does justice to the evidence about pragmatic reasoning), is that it must model the reasoning as a single, overall process, encompassing everything from "low-level" disambiguation to "high-level" inferences about relationships among utterances in discourse, and embracing both reasoning about language and reasoning about the world. In Section 2.5, we show how the computational model of interpretation as abduction meets this holistic challenge. This background sets the stage for our subsequent discussion—a mix of analogy, analysis and implementation that, we think, shows why our account is not just a conceivable view but an inevitable consequence of the kind of collaboration interlocutors undertake in conversation.

2.1 Meaningfulness

We talk about the world around us, and the grammar allows us to use information in the conversational situation to derive the content of an utterance on a particular occasion. Utterances containing demonstratives, as described by Kaplan (Kaplan, 1989), offer an illustrative example. Suppose a speaker, call him Andrew, points at a pot on the stove in which potatoes are simmering, and presents his interlocutor Bess with an utterance of (4).

(4) Those are heirlooms.

He thereby means that **p**, those potatoes, the very objects he has pointed at, are **heirlooms** (namely, of a plant variety developed before 1950 for attributes like hardiness and flavor suitable to historic small-scale pesticide-free farming). Thus, the information that Andrew expresses with this utterance explicitly concerns these individuals from the conversational situation.

To plan and recognize such utterances, language users rely on the same knowledge of language across all situations. Words like *those* have a consistent linguistic meaning that underwrites the content they contribute on different occasions. The content of *those* requires an act of demonstration whose object is a plurality; when this requirement is met, this plurality serves as the semantic value of *those*, and compositional semantic rules determine how this value contributes to the content of the uttered sentence.

For the discussion that follows, it will be useful to indicate how an explicit theory could state such generalizations about (decontextualized) meaning and content. Grossly simplifying and specializing the relevant syntactic and semantic principles, we might spell out the theory required to link utterance (4) to its content as in (5).

(5) a. Any utterance of the expression *heirlooms* is an indefinite plural NP and refers to the property **heirlooms**.

b. If uttered in association with an accompanying demonstration that designates the objects **S**, then the definite plural NP *those* refers to **S**.

c. If E_1 is a definite plural NP that refers to **U** and E_2 is an indefinite plural NP that refers to **P**, and E_1 and E_2 are part of a larger utterance of E_1 *are* E_2, then this larger utterance is a sentence and expresses the proposition that objects **U** have the property **P**.

The theory in (5) has as a logical consequence the semantic characterization of (4) given in (6):

(6) If uttered in association with accompanying demonstration that designates the objects **S**, then *those are heirlooms* expresses the information that objects **S** have the property **heirlooms**.

Note how the formal structure of (6) represents the link between meaning and content parametrically. The meaning of the utterance includes a free parameter **S**. The grammar constrains the value of **S** based on the associated demonstration and the objects it designates in the conversational situation. We obtain the content of Andrew's utterance from its meaning by instantiating **S** to **p**, taking into account the fact that Andrew's demonstration in fact indicates **p**. Thus we derive the specific content that (4) expresses the information that **p** have the property **heirlooms**.

A key question in framing our intuitions about meaningfulness is to sharpen our understanding of the constraints and processes that link parameters and values in such cases. We assume that the reference of the demonstration is DETERMINED by the speaker's intentions. Even if the speaker's act of pointing is poorly aimed, and the hearer is plainly unable to see this act, under normal circumstances the demonstration indicates the potatoes if the speaker intends it to. On this story, your intention in demonstrating an object determines what you refer to with a diectic pronoun in much the same way that your intention to go to your refrigerator determines your destination. Within wide cognitive limits, you can choose any destination you want—but the success of the choice depends in part on the wider world in which the choice is made. (You will have no destination if the refrigerator has been destroyed. You will have a destination if, unknown to you, the refrigerator belongs to someone else—but it will not be your refrigerator.) Thus, for an utterance of (4) to be MEANINGFUL, it just needs the designated objects to flesh out its underspecified meaning to a specified proposition. The resolution of this underspecification follows from the facts about the communicative action and the situation in which it is performed.

Reasonable intentions are constrained by their place in plans, and in particular by the goals that they serve. A demonstration that does not contribute to the conversational goals may determine a reference, but it is still fundamentally flawed. And even quite reasonable intentions can fail if the circumstances don't cooperate. For example, we may be ignorant or misinformed about meaning in conversation. Take (4), and imagine that Andrew has pointed to a covered pot, designating its contents. His interlocutor Bess may not know what's cooking. It might be the potatoes bought at the farmer's market this morning; it might be the beets. One way to describe this is that Bess incompletely identifies what Andrew means. Andrew in fact means that the potatoes are heirlooms. That's what's in the pot. But Bess doesn't know whether Andrew means the potatoes are heirlooms or means the beets are heirlooms. If Andrew really wants to inform Bess that the potatoes are heirlooms, this communicative action will not do it. In this sense, his intention is flawed.

This uncertainty can infect the speaker as well as the hearer. Suppose that Andrew himself uses (4) without knowing whether the pot contains the potatoes or the beets.

Again, we can defend the intuitive picture that Andrew still does mean that the potatoes are heirlooms. That's what he's pointed at. Still, he has expressed a proposition that he himself cannot identify. Indeed, in general, it seems that we cannot assign public meaning to utterances—as our intuitions invite us to do—without recognizing cases where speakers and hearers are ignorant of crucial matters of meaning (Kripke, 1972; Burge, 1973; Putnam, 1975). These intuitions are broadly accepted in the literature in the Philosophy of Language. We would like a pragmatic theory that endorses and explains them.

2.2 Coordination

At the same time, we take seriously arguments that mutuality of information is decisive in pragmatic reasoning. The importance of mutuality is supported by two sorts of theoretical and (loosely) empirical considerations. The theoretical reasons were developed by David Lewis (Lewis, 1969), and apply generally to cases requiring the coordination of actions over a population of rational distributed agents. Using the theory of cooperative games, Lewis shows that the adoption of an ideally rational, CONVENTIONAL signalling system requires mutual knowledge.[2] Lewis' insights have become the basis of theories of economic transactions (Aumann, 1976; Geanakopolos, 1994) and communication (Fagin, Halpern, Moses & Vardi, 1995). These theories reinforce Lewis' arguments that rational cooperation is impossible in many cases without mutuality.

Similar theoretical considerations have made their way into computationally inspired theories of discourse (Perrault & Allen, 1980; Joshi, 1982). Well known reasoning problems like the muddy children puzzle, although artificial in some ways, can be made to seem plausible and compelling, and depend essentially on the assumption that a public announcement to a group creates mutual knowledge among the members of the group.

Also, it is generally accepted that a distinction between given and new information is important in understanding many linguistic phenomena, and there are many examples in the literature seeming to show that mutual knowledge captures the appropriate notion of mutuality in many cases. One particularly convincing and well presented example of this sort is the "Monkey Business" scenario in (Clark & Marshall, 1981), showing that a felicitous utterance of *Have you seen the movie playing at the Roxy* requires the speaker to believe that the reference of *the movie playing at the Roxy* is mutually known.

Such examples are especially natural and convincing when one takes into consideration only the speaker's beliefs about the hearer's attitudes: you should not ask someone to hand you the crescent wrench unless you expect that they know what a crescent wrench looks like, and can see the wrench. But true mutuality requires arbitrary iterations of attitudes; for instance, a and b do not mutually know P unless a knows that b knows that a knows that b knows that P. Common-sense reasoning tends to break down over such iterations, and it takes a certain amount of work even to construct examples involving more than two iterations. This makes it difficult to construct a compelling empirical case for the mutuality of given information in discourse. It is the combination of simple examples with the difficulty of finding any principled place to cut off the need for iterations,

[2]We use the terms 'knowledge' and 'belief' here, depending on which tends to sound more natural. But the relevant attitude needn't be knowledge or belief. It could be supposition, or any attitude that serves to guide actions in a given, circumscribed situation.

and with the importance of mutuality in economic theories that makes mutuality such an attractive concept in the theory of discourse. But the deployment of this assumption in discourse has to be qualified in the same way that it is in economics; this is an assumption about ideal rationality, that has to be applied with care to human reasoning in realistic situations.

Some authors, such as Sperber & Wilson (Sperber & Wilson, 1995, pp. 15–21) have claimed that, since a mutual belief has infinitely many logically independent consequences, we cannot have mutual beliefs. If we took such arguments seriously, we would have to conclude that no observations could lead us to conclude that we have ten fingers. Nevertheless, if we require mutuality in discourse, even as an idealizing assumption, we need to show how it is possible to reason plausibly to conclusions about mutuality, and maintain mutuality in the course of a dialogue. We believe that this can be done. For an account of mutuality in terms of default suppositions, see Thomason (Thomason, 2000).

2.3 Reconciling these intuitions

Approaches deriving from Stalnaker's publications on presupposition tend to share the idea that presupposition is primarily attributed to speakers (not to sentences) and that it is related to the common ground, a mutual attitude that tracks what has been established in a conversation. One form of the idea, introduced in (Stalnaker, 1975) and advocated in some of Stalnaker's later works, is that a speaker's presuppositions are the things she believes belong to the common ground. We reject this idea.

Instead, we work with a more general notion of presupposition deriving from the structure of intentions, which we take to involve not only a GOAL (a state of affairs to be achieved), but a PLAN (a partially specified way of achieving the goal), and PRECONDITIONS, or ways that the world is assumed to be, and on which the success of the intention depends. Communicative intentions, like other intentions, can have preconditions. For example, as described in Section 2.1, the grammar may encode preconditions about the conversational situation that are required for an utterance to be meaningful. Accordingly, a speaker's communicative intention in using that utterance must take on the commitment that these preconditions are met. It is these preconditions that we identify with speaker presuppositions, or the presuppositions of an utterance. While we acknowledge that speakers will track the common ground to ensure their intentions are recognized—clearly an intention that cannot be recognized is flawed—we do not go the further step of assuming that the speaker must take these preconditions to be part of the common ground.

Analyzing presupposition via intention is similar in some ways to an alternative account of speaker presupposition that Stalnaker mentions (Stalnaker, 1973), according to which a speaker presupposes the things she is disposed to take for granted.[3] This alternative is not as prominent in Stalnaker's work or as influential as the one that appeals to common ground, but it has the advantage of avoiding the problem of informative presupposition. As far as we know, clarifying what it is to "take for granted" in terms of intention is a new idea. This characterization has the advantage of relating presuppo-

[3]Stalnaker's views on presupposition have evolved over a long period of time, and are complex. For a readable and informative discussion of this topic, see Simons (Simons, 2003).

sition to what is meant through the idea that meanings are intentions of a certain sort. And—most important, for our purposes—it provides insights into the related reasoning processes, since the intention that stands behind an action of any kind will explain the action by providing a goal, a method of achieving that goal, and certain preconditions. This explanation will involve knowledge of the same sort that is needed for planning: knowledge of the effects and preconditions of actions. It is through formalizing this reasoning, and the knowledge it depends on, that we are able to reconcile our view of presupposition with the requirement, explored in Section 2.2, that interlocutors coordinate using mutually-supposed information.

Treating informative presupposition, as we do, as a matter of "enlightened update" eliminates the need to explain the main body of phenomena either in terms of repair strategies arising from the violation of a rule (Lewis, 1979), or in terms of pretense (Clark, 1992), or in terms of sequential updates of the common ground in which presuppositions are added before the primary assertion (von Fintel, 2000). All of these ideas have some plausibility, and may prove to be useful parts of pragmatics. But the first two of them, we feel, are not appropriate for a phenomenon that is as common and routine as informative presupposition. And we know of no independent way of motivating multi-stage updates directly from considerations having to do with pragmatic reasoning. Without motivation of that kind, it looks very much like an *ad hoc* solution, whose only purpose is to save an account of speaker presupposition that requires presuppositions to belong to the common ground.

In (Stalnaker, 2002), Stalnaker suggests a view of accommodation as a side-effect of conversants' attempts to cooperatively align their presuppositions in the course of a conversation. This is very close to the ideas we discuss in Section 4.4, but again, we are able to relate this to an account of the reasoning that is much more detailed.

In fact, our insistence on a model of the reasoning that is sufficiently detailed to support an implementation is perhaps the most important difference between our approach and others in the literature, since it shapes our position in fundamental ways. This methodology requires us to choose a reasoning domain, and to deal systematically with the cases of presupposition accommodation that arise there. Since tractable reasoning domains tend to be relatively simple, though they provide many examples of presupposition accommodation, these examples will seem relatively simple and routine. We are betting that a methodology working with such examples, but accounting for the reasoning with unprecedented detail, will be more successful in pragmatics than a more traditional linguistic methodology involving a wide range of more complex constructed examples.

2.4 Pragmatics and cognitive modularity

Linguistic methodology favors an approach that leads from intuitions about linguistic examples to structural generalizations, which then are explained in terms that are (1) modular and (2) independent of the reasoning mechanisms. Thus, a grammar of a language is a knowledge-level account of the language's structures. The grammar is supported by reasoning procedures—for instance, the procedures involved in parsing—but there is no reason to expect these procedures to be similar to and continuous with the procedures that are involved in general-purpose common-sense reasoning. Linguists do not in general

seek to integrate their theories with knowledge-level accounts of nonlinguistic domains, and usually modularize their grammars into more or less independent subsystems.

These practices are not well adapted to pragmatics. Task-based discourse is closely connected to the domain reasoning associated with the task. Pragmatic phenomena such as reference resolution, disambiguation, implicature detection, and the identification of discourse relations depend heavily on domain knowledge. And pragmatic reasoning is often similar or even identical to domain reasoning. Suppose Bess asks Andy, who is standing by, for the salt. If Bess is eating, Andy correctly infers that Bess wants the small salt shaker. If Bess is cooking and holding a tablespoon measure, Andy correctly infers that Bess wants the box of salt. Andy's reasoning here is identical to or similar to the reasoning he employs in planning how to use salt for various purposes. Examples such as this are commonplace.

We can't hope, then, to isolate pragmatic reasoning from domain reasoning. Neither can we assume that pragmatic reasoning divides into disconnected special purpose modules, devoted to different pragmatic phenomena, because the preferences that govern the selection of a favored interpretation are OVERALL preferences. The following discourse illustrates the point.

(7) Ann thought she might be able to get help with her computer problems from Betty. She is a systems analyst.

The fact that *Ann* is the subject of the first sentence creates a *prima facie* preference for Ann as the reference of *she* in the second sentence. However, the two sentences can be regarded as an explained/explanation pair if *she* is resolved to *Betty*, and a plausible implicature—that Ann thought that a systems analyst would be helpful with her problems—would also be generated. These factors create an overall preference for Betty as the reference of *she*. Many similar examples can be found in (Hobbs et al., 1993).

Thus, the reasoning associated with various pragmatic domains (anaphora, reference resolution, metonymy resolution, implicature, presupposition, etc.) must produce commensurable preferences, and must access and overlap with domain reasoning. Under these circumstances, we believe that the best working hypothesis is to account for the reasoning in as nonmodular a way as possible: pragmatic reasoning uses knowledge structures and reasoning that are similar to those used in domain reasoning, and is holistic, dealing at once with all pragmatic phenomena.

Also, exploiting a theme from Grice, who in (Grice, 1957) assimilated discourse meaning to cases involving language and nonlinguistic cases involving only cooperation and intention recognition, we believe that pragmatic reasoning will have much in common with the reasoning that supports collaborative human activity.

2.5 Interpretation as abductive intention recognition

Abduction provides a reasoning framework for pragmatics that is properly general-purpose, and that supports domain reasoning such as explanation and planning. The knowledge structures are intentions and their ingredients, and preferences over these intentions. The framework inherits the naturalness and plausibility of intention-based

approaches to pragmatics, and with some natural extensions concerning helpfulness and the maintenance of mutual information, it can also be used to support collaboration.

The logic programming approach developed by Hobbs and his associates to implement this knowledge-level picture of interpretation as abductive explanation of the utterance provides an equally appealing account of the reasoning. Hobbs took proving logical form of the uttered sentence to be the goal of the hearer's reasoning,[4] in effect taking the hearer to be asking the question "What does this sentence mean?" We prefer to think of the goal as proving that a certain intention supports the uttered sentence; in effect, the hearer is always asking "What did the speaker intend by saying that?"

As we said, an abductive proof will in general require new assumptions, which are added to the conversational agent's stock of working assumptions. This provides a way of distinguishing the new information in an utterance (the assumptions created by the process of interpretation) from old information (the stock of working suppositions that are used as background for the interpretation). The distinction is similar to the one deriving from Stalnaker between the pragmatic presuppositions of an utterance and its new information, and it is formulated in an implementable reasoning framework that combines grammatical requirements for wellformedness and meaningfulness with the informational demands of domain reasoning and conversational coordination. Thus, Hobbs's abductive approach shows how the grammatical requirements for meaningfulness and informational demands of coordination can be integrated in conversational reasoning.

Thinking of interpretation as the derivation of an INTENTION rather than of a sentence meaning introduces a more subtle set of distinctions into pragmatic theory. An intention can introduce propositional assumptions in various ways. If, for instance, the intention is assertional, it will have a primary goal of updating the common ground with a proposition. But it may also have preconditions—propositions that are required for the feasibility or appropriateness of the intention—which can be new information, even though they are not part of the content of the assertion.

The present paper, like our earlier ones (Stone & Thomason, 2002; Stone & Thomason, 2003), is an attempt to put these different ingredients together.

3 Collaboration and conversation

As we explained in Section 2.4, we subscribe to the methodological principle that pragmatic reasoning is continuous with domain reasoning. We now expand on this by developing the idea that reasoning about conversational intentions parallels reasoning about intentions in other kinds of collaborative activity.

In fact, the collaborations that we undertake as embodied agents in the physical world, because they lack the abstraction of linguistic interaction, often offer the best illustration of pragmatic principles, and of the role of abductive intention recognition in the reasoning.

We wish to emphasize the following points concerning intentional collaborative activity:

[4] Hobbs used a somewhat idiosyncratic version of logical form, with some similarities to linguistic LF.

(1) Any intentional activity, collaborative or not, requires us to make private commitments about how the world is.

(2) Coordinated activity requires keeping track of mutual information, and distinguishing it from what is private.

(3) Mutual information must be updated and maintained, along with private beliefs.

(4) Agents capable of intention recognition can transmit information by performing actions in public.

(5) If the mechanisms of intention recognition are mutual, public actions can create new mutual information.

(6) In particular, an agent's public performance of an action that is mutually known to require a commitment C for its successful performance will add to the mutual information the proposition that the agent believes C. Normally, this will result in the addition of C to the stock of mutual information.

Characteristics of Intentional Collaboration

Stalnaker (Stalnaker, 1981) characterized assertional acts as speech acts that add a proposition to the common ground. But assertion is not the only method of incrementing the common ground: the mechanism described in Characteristic (5) provides another way, based on intention recognition, for intelligent collaborating agents to add private commitments that can be mutually inferred from shared information (including mutually observed actions) to the common informational background. Updates of this kind do not require verbal communication. Nor are they meant, in the sense of (Grice, 1957). In the cooking example below, Andrew does not mean by mixing the spinach in the salad that he has washed it; he does not intend Bess to recognize that he has washed the spinach in part because of her recognition of his intention for her to recognize this.

However, we can supplement this mechanism with explicit communication; we can and sometimes must provide verbal information to accompany what we do, so that our private commitments are correctly recognized.

We illustrate these points with an extended example. Andrew and Bess are cooking. Bess asks Andrew to wash and dry some spinach for a salad. Bess then turns her back to Andrew and attends to a pot on the stove. It is boiling ferociously, and she can neither see nor hear what Andrew does. Andrew cleans the spinach. Bess turns to face Andrew again.

In this situation, we have to distinguish between the state of the collaborative activity, the agents' private knowledge, and their mutual information. In the current state, Andrew has cleaned the spinach and knows that he has done so. But Bess does not know whether Andrew has cleaned the spinach. And we can suppose that Bess and Andrew both mutually recognize that Bess does not know this. So it is mutually recognized at this point that maybe Andrew has cleaned the spinach but maybe not.

Let us continue the example. In plain view, Andrew mixes the spinach with the other salad ingredients. In fact, this action furthers the joint activity in a straightforward way; Andrew and Bess are one step closer to a good salad.

Consider how Bess updates her information as she watches this. For all she knows, Andrew may not yet have washed the spinach. But on seeing Andrew add the spinach to the salad, Bess can rule out this possibility. Andrew himself is in a position to know whether he has washed the spinach, and is committed to making a good salad. So if he hadn't washed the spinach, he wouldn't have added it to the salad. Andrew's action thus eliminates the ambiguity in the state of the activity. Afterwards, because of this ENLIGHTENED UPDATE, Andrew and Bess enjoy the mutual information that the spinach is washed and clean, and so the meal preparation continues smoothly with this as common ground.

Andrew's action of mixing the salad was public, and its performance was mutual to the participants. And Bess's inference about the tacit action used mutual information about salad making. So Bess can add the tacit action not only to her private store of information, but to the mutual common ground.

In this setting, Andrew's action of cleaning the spinach is what we call a TACIT ACTION. An action in a collaboration is tacit if some of the collaborators do not know whether it has taken place or whether it has had the appropriate effects. We will call other actions PUBLIC ACTIONS; collaborators know when public actions occur and whether they succeed.

Tacit actions give us clear grounds to distinguish between the state of the activity and participants' information—they create a gap between what participants have done and what they mutually know. Now, we do not claim that all such gaps are due to tacit actions. There may be exogenous events that some collaborators can observe, but some cannot. For example, perhaps only Bess will see when the contents of her boiling pot have cooked. This would be a tacit event, not a tacit action. Andrew might still infer that this event has taken place from observing Bess's actions—say, by watching Bess turn off the heat or empty the pot. Moreover, there may be uncertainties about the state of the collaboration that aren't usefully traced back to any particular event in the collaboration at all. For example, perhaps it's Bess's kitchen and so only she knows that the small ceramic bowl on the countertop is a salt cellar. Andrew might discover this by observing Bess adding a pinch of white crystals from the bowl to the pot on the stove. Thus, Andrew's reasoning in explaining Bess's actions may very well be similar whether he is resolving uncertainty that arises from tacit actions in the collaboration, from unobserved exogenous events, or simply from incomplete background knowledge. Nevertheless, we focus on the case of tacit actions in this paper. They are a clear case, they are an indispensable ingredient of problem-solving interactions, and they have, we think, a particular significance for conversation.

In this paper, we explore the agents' reasoning in response to a mix of public and tacit action in a collaboration. When a public action follows some tacit actions, interlocutors can fill in gaps in their knowledge by exploiting their understanding of the ongoing collaboration. Whenever a public action is taken, the collaborators all discover that the state of the activity already made the public action appropriate to the state in which that action was taken. If the public action was not appropriate, its agent should not have so acted.

The collaborators can thus jointly infer that the tacit actions must have occurred and, in relevant respects, succeeded. As a result, collaborators can make an enlightened update to their information state—they can proceed in effect as though they always knew about the tacit actions and their relevant effects.

Tacit actions, and the distinction between mutual and private information, are regularly taken into account in reasoning about collaboration and conversation. In particular, mutual beliefs are crucial in effective collaboration, and as Herbert Clark and others have noted, speakers are careful to maintain mutuality. Suppose that at this point in the cooking activity, it is mutually known that a pinch of cayenne and a tablespoon of cumin need to be added to the pot. The spice containers are not labeled, but Bess knows which is which. She hands a container to Andrew. Given mutual information about the task, Andrew can infer that he is to add the spice to the pot. But Bess's previous contribution was flawed, since at this point Andrew doesn't know which spice it is and so doesn't know how much to add. By keeping track of shared information, Bess can infer that simply handing the spice to Andrew would be uncooperative. Repairing this flaw could lead to disambiguation by adding a verbal clue—say, by saying "here's the cumin" as she passes the spice container.

Of course, people engaged in cooperative activity can make mistakes, and produce ambiguous and even misleading actions, just as they sometimes produce ambiguous and even misleading utterances. But the fact that we classify these cases as mistakes shows that we recognize the importance of tracking mutual information in these cases.

If Bess is a worrisome type, or feels that Andrew is less than conscientious about washing vegetables, she may ask Andrew for explicit confirmation that he indeed washed the spinach. But if this is a task they have performed many times, and Bess has no worries about Andrew's competence, the enlightened update is added to the common ground, just as if the action had been verbally confirmed or publicly observed. Bess can say things like

(8) Did you wash the parsley when you washed the spinach?

for instance, with perfect propriety. In collaborative activity, enlightened update not only serves as a valid way to add information to the common ground, but it is frequent and pervasive.

4 Pragmatic reasoning in a cooperative planning domain

In this section, we flesh out our case by presenting a simple model of the coordinated reasoning of collaborators in conversation. As in our previous papers (Stone & Thomason, 2003; Stone, 2004b), we consider two-party interactions, and conceptualize the generation process of the one agent S and the understanding process of the other agent H in symmetrical ways. The reasoning problem that grows out of the discussion of Section 3 is how to handle the uncertainty about the state of the ongoing task that arises because of the collaborators' incomplete knowledge about the actions that have been performed. The thought experiments there suggest that language users may face other sources of uncertainty in everyday face-to-face interaction; we leave this for future research.

Concurrently, we describe how we have implemented this model in a conversational agent, COREF, that works together with its interlocutor to identify objects from a scene displayed in a graphical computer interface. Our demonstration system plays a referential communication game, much like the one that pairs of human subjects play in the experiments of Clark and Wilkes-Gibbs (Clark & Wilkes-Gibbs, 1986). Two subjects are presented with a display containing the same set of objects. One, the DIRECTOR, sees them arranged in a specific linear sequence, while the other, the MATCHER, sees them shuffled arbitrarily across their visual space. The interlocutors must go through the objects one by one so that the matcher lays them out in the director's order.

We describe each episode in this game as an activity involving the coordinated action of the two participants. The director D knows the referent R of a target variable T and the matcher M needs to identify R. COREF can play either role, D or M, using the objects in the graphical display as candidate targets and distractors, and using text as its input and output.

We exemplify what COREF does in (9). The corresponding graphical display is shown in Figure 1.

(9) C_1:This one is a square.
U_2:Um-hm.
C_3:It's light brown.
U_4:You mean like tan?
C_5:Yeah.
C_6:It's solid.
U_7:[Moves the object into position].

Here COREF (C) and the user (U) exchange six utterances—and one action—in the course of identifying the tan solid square that fills the next square on the game.

COREF uses the same task knowledge and the same grammar whichever role it plays. It reasons from a specification of collaborative reference that includes tacit actions, and thus makes frequent recourse to enlightened updates. These updates figure not only in utterance understanding and generation, but also in planning and recognizing nonlinguistic public actions, such as the user-interface actions represented by U_7. Of course, COREF also draws on private knowledge to decide how best to carry out its role; for now it describes objects using the domain-specific iteration proposed by Dale and Reiter (Dale & Reiter, 1995). The knowledge we have formalized is targeted to a proof-of-concept implementation, but we see no methodological obstacle in adding to the system's resources.

4.1 Formalizing context in collaborative activity

We model a collaborative planning activity as a dynamical system in which the states of a domain are altered by successive actions. Each point in the activity is characterized by a state s of the domain, and by the knowledge the agents have of the domain state and of each others' knowledge. The epistemic aspects of such systems can be modeled using possible-worlds semantics; see Fagin et al. (Fagin, Halpern, Moses & Vardi, 1995). In (Reiter, 2001), Raymond Reiter shows how to incorporate these ideas in settings of the sort used in Artificial Intelligence to formalize planning.

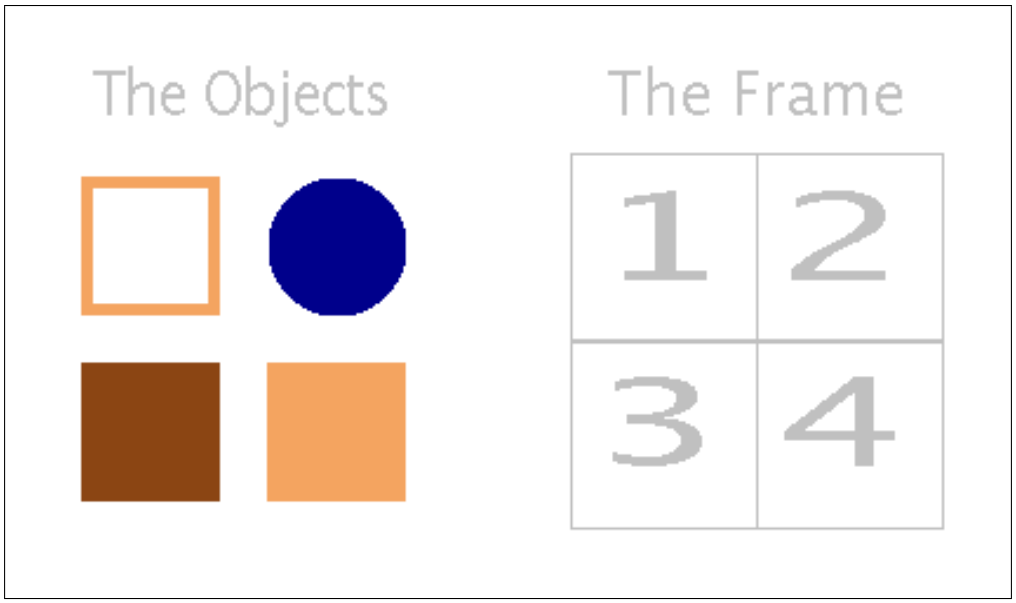


Figure 1: The matcher's view of a collaborative reference task. An array of objects is displayed in an unordered array at left. These objects need to be placed, as specified by the director, into the frame at right.

To model collaborative conversation, in particular, we take each state to specify an abstract, objective body of information that constitutes the conversational situation—the information state (IS) of the conversation (Poesio & Traum, 1997; Larsson & Traum, 2000). Thus, it will spell out the attributes of dialogue state that are familiar from computational theories of discourse, including: sets of propositions contributed to the conversational record (Stalnaker, 1981), plans and problem-solving activities that are underway (Lochbaum, 1998), outstanding interlocutor obligations (Traum & Allen, 1994), linguistic forms of prior utterances (Purver, 2004), etc.

Precisely what information goes into the information state depends on what interlocutors are doing. For a task of collaborative reference, the IS tracks how interlocutors together set up and solve a constraint-satisfaction problem to identify a target object. In any point in this problem-solving, D and M have agreed on a target variable T, a set of candidate referents C that could serve as the value of T and a set of constraints that the value of T must satisfy. We call this body of information a CONSTRAINT NETWORK (Mackworth, 1987). When M recognizes that these constraints identify the target referent R, the task ends successfully. Until then, D can take actions that contribute new constraints on R. Importantly, what D says adds to what is already agreed about R, so that the identification of R can be accomplished across multiple sentences with heterogeneous syntactic structure.

In COREF's domain, the overall game is to identify a set of objects in turn. Part of the information state—the task state—tracks this activity in terms of episodes of collaborative reference. We represent the task state through a tuple $c^* = \langle U, O, P, T, C \rangle$, where U is the overall universe of discourse (a set of properties and objects), O is the set of referents

yet to be identified, P is a set of agreed propositions, T is a stack of tasks (where each task specifies a joint goal that the collaborators are engaged in and the actions they can take in support of that goal), and C is a set of constraint networks (one for each target referent). In this state, U depends directly on the visual display; O starts out containing all the objects on the display but is narrowed down as these objects are identified one by one. P straightforwardly accumulates the information that interlocutors have exchanged. Meanwhile, T and C evolve in a richly structured way according to interlocutors' actions.

This model contrasts with traditional plan-based models, as exemplified by Heeman and Hirst's model of goals and beliefs in collaborative reference (Heeman & Hirst, 1995). Our representations are simpler than Heeman and Hirst's but support more flexible dialogue. For example, their approach to (9) would have interlocutors coordinating on goals and beliefs about a syntactic representation for *the tan solid square*; for us, this description and the interlocutors' commitment to it are abstract results of the underlying collaborative activity. We formalize the activity as a cumulative and open-ended collaborative process, much as advocated by Rich et al. (Rich et al., 2001).

Our information state tracks other features of the conversational interaction in addition to domain activity. For example, our IS also tracks the salience of entity and property referents and explicitly stores the previous utterance, enabling subsequent reference to it. Meanwhile, the IS also allows subtasks of questioning or clarification that interlocutors can use to maintain alignment. In fact, the same constraint-satisfaction model is used not only for identifying displayed objects as part of the collaborative reference game but also for identifying abstract entities, such as properties, in pursuing clarification subdialogues. The key antecedent for these extensions is Purver's (Purver, 2004) characterization of clarification of names for objects and properties. We extend Purver's work to allow a cumulative and open-ended treatment of clarification; when we describe things, our descriptions grow incrementally and can specify as much detail as needed.

4.2 Action, tacit action, uncertainty and coordination

The collaborative context evolves over the course of the dialogue through the domain-dependent set of action types, $\mathcal{A}$, that interlocutors can take. Each action a that interlocutors take has the form $\sigma(\alpha)$ where $\alpha \in \mathcal{A}$ and σ instantiates the free parameters of α. Doing a effects a *deterministic* transformation on the current context. We can formalize this by way of an update function that recalls the notation of the situation calculus:

$$s_{t+1} = \mathsf{do}(a, s_t) \tag{1}$$

(Again, see Reiter (Reiter, 2001).) Thus, in the deterministic case, the state of the domain can be characterized either by specifying the state itself or by specifying the sequence of actions that led to it. Such a sequence will produce a unique state from the initial state of the domain.

In our formalization, we assume that each action type is either PUBLIC, which means that any time any action of this type is performed, its occurrence becomes mutual knowledge among all agents, or TACIT, which means that only the agent that performs an action of this type knows about its occurrence. In other words, our model embodies the simplification that observability is just a matter of the type of action performed, idealizing away

from its dependence on the contingent relationships among the agents.

As agents perform tacit actions, they must coordinate with one another to maintain mutual knowledge of the domain state. With COREF, we develop a precise account of one way this coordination might proceed by regimenting how the collaboration takes place. In particular, we imagine that the collaboration intersperses stages of uncertainty with CHECKPOINTS—states in which both agents have full, mutual knowledge of the domain state. We assume that agents achieve this by acting in conformity to a Principle of Coordination Maintenance (stated formally below). In brief, according to this principle, agents will limit what they do and elaborate how they present it so that they can keep in synch with one another.

Suppose that in the initial state of the entire system, both agents have mutual knowledge of the domain state. At this point one agent, S, performs a sequence of tacit actions. H cannot observe these actions, but can assume that S is acting in pursuit of their joint collaboration. In the subsequent state of the system (1) the domain is in the state that results from performing these actions, (2) S knows that the domain is in this state, and knows that H considers any legitimate sequence of actions performable by S on the initial domain state to be possible, and (3) H knows that S knows the state of the domain, but as far as the domain is concerned knows only that it has reached a state that can be obtained by a legitimate sequence of S's actions. This space of possibilities can be represented as the set of branches over a tree whose nodes are actions. Under our assumptions, both S and H have access this space. However, whereas S knows what point in this space they actually occupy, H lacks this information. We call this space THE HORIZON; if c indicates the checkpoint state, $Z(c)$ indicates the corresponding horizon.

After performing a sequence of tacit actions, S chooses an observable action e. This action will advance the joint activity, but crucially, it also has to coordinate S and H by resulting in a checkpoint.

In understanding e, H's task is to identify the transition that e effects: what the state of the activity was beforehand, what e itself does, and what the state of the activity is afterwards. This amounts to recognizing the sequence of tacit actions preceding e, and how e effects this context, which can be done by understanding S's intention in undertaking e, including the private commitments of this intention. S's task in generation is to produce an action that enables H to perform this recognition. Since H is using abductive reasoning, this means that S must produce an observable action whose preferred explanation is the intention that actually underlies this action.

In COREF, the action set $\mathcal{A}^*$ has observable task actions including acknowledging a prior contribution (notated by the symbol `ack`); adding a constraint C to the constraint network for a target t (notated `addcr`(t, C)); placing a specified object x at the next point in the desired sequence (notated `place`(x)); inquiring whether some other action a can now be taken (notated `ynq`(a)); and rejecting an action a as impossible (notated `no`(a)). In addition, much of the flow of the activity is modeled through tacit domain actions, including actions of selecting the referent sequence, initiating collaborative reference to a particular target referent t (notated `Push(CollabRef(`t`))`), marking the target t as identified (notated `Pop(CollabRef(`t`))`), triggering a clarification subtask by marking an utterance u as incompletely understood (notated `Downdate(`u`)` following Purver (Purver,

2004)), or reinstating the effect of an utterance once its meaning has been agreed through clarification (notated `Reinstate(`u`)`). These moves allow COREF to handle clarification, grounding, and other kinds of implicit task progress with a flexible, declarative model and consistent processes of interpretation and generation, based on enlightened update.

In (9) for example, the system represents its moves as successively constraining the shape, color and pattern of the target object. The user's clarification request at U_4 of (9) marks the agent's description of color as problematic and so triggers a nested instance of the collaborative reference task. At C_5 of (9) the system adds the user's proposed constraint. With C_6, the system assumes the nested clarification task is solved and that the color constraint is reinstated, and then continues the main task by identifying the pattern of the target object. The user's action of moving the object into position at U_7 shows that they have tacitly identified the object and are continuing on with the task.

The full analysis of task actions in dialogue (9) is presented in Figure 2. To take one case, in interpreting the user's utterance *You mean like tan?* (U_4 of (9)), COREF interprets the user as signaling the following sequence:

(10)
1. Downdate the effects of the problematic utterance C_3.
2. Initiate clarification.
3. Start collaborative reference, targeting COREF's intended property S associated with *tan*.
4. Start a question-answer task in which COREF is obligated to answer a user question.
5. Ask the question: `ynq(addcr(`S`,` $S =$ `sandybrown)).`

Only the last two of these can be seen as directly associated with the semantics of the utterance. The others are tacit actions that are recognized abductively as part of COREF's recognition of a communicative intention that makes sense of the user's utterance.

COREF uses its action representation not just for domain problem-solving but also to manage aspects of dialogue such as the attentional state. The grammar, for example, specifies that each subject noun phrase updates the context to give its referent the status in the givenness hierarchy of being IN FOCUS (Gundel, Hedberg & Zacharski, 1993). This update makes the referent available as a preferred referent for subsequent pronouns. In (Stone & Thomason, 2003) we explain in detail how we represent such transitions in the framework of interpretation as abduction.

Many simplifying assumptions are incorporated in this model. The domain representation employs a highly specific formalization of the communication topic, and of the content that needs to be communicated at any given point. In this respect, our domain is much more regimented than ordinary, informal conversation. We also assume that actions are asynchronous, and follow the rather rigid protocol described above.

We are, in effect, following the methodology of Artificial Intelligence, in investigating a complex reasoning problem by investigating how it plays out in a simplified, formalizable special case. Despite the simplifications, we believe that the model is realistic enough to shed light on some of the important problems of pragmatics, and that it can be extended incrementally to cover more complex cases.

Utterance	Task Moves	Resulting Task State	
		Task Stack	Facts
		ClarkGame	square(a) solid(a) sandybrown(a)
C: *This one is a square.*	(tacit) Push(Collabref(t_1)) addcr(t_1,square(t_1))	CollabRef(t_1) ClarkGame	square(a) solid(a) sandybrown(a) square(t_1)
U: *Um-hm.*	ack	CollabRef(t_1) ClarkGame	square(a) solid(a) sandybrown(a) square(t_1)
C: *It's light brown.*	addcr(t_1,sandybrown(t_1))	CollabRef(t_1) ClarkGame	square(a) solid(a) sandybrown(a) square(t_1) sandybrown(t_1)
U: *You mean like tan?*	(tacit) Downdate(C_3) (tacit) Push(Clarify(C$_3$,S_1)) (tacit) Push(CollabRef(S_1)) (tacit) Push(YNQ) ynq(addcr(S_1, $S =$ sandybrown))	YNQ CollabRef(S_1) Clarify(C_3,S_1) CollabRef(t_1) ClarkGame	square(a) solid(a) sandybrown(a) square(t_1)
C: *Yeah.*	addcr(S_1,$S =$ sandybrown)	CollabRef(S_1) Clarify(C_3,S_1) CollabRef(t_1) ClarkGame	square(a) solid(a) sandybrown(a) square(t_1) $S =$ sandybrown
C: *It's solid.*	(tacit) Pop(CollabRef(S_1)) (tacit) Pop(Clarify(C$_3$,S_1)) (tacit) Reinstate(C_3) addcr(t_1,solid(t_1))	CollabRef(t_1) ClarkGame	square(a) solid(a) sandybrown(a) square(t_1) $S =$ sandybrown sandybrown(t_1) solid(t_1)
U: *[Action]*	(tacit) Pop(CollabRef(t_1)) place(a)	ClarkGame	square(a) solid(a) sandybrown(a) square(t_1) $S =$ sandybrown sandybrown(t_1) solid(t_1)

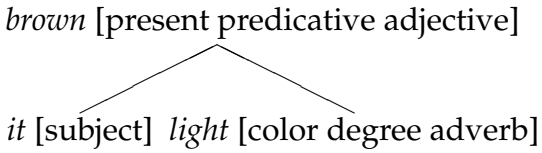
Figure 2: A trace of task actions and task states through which COREF represents and participates in the interaction of (9).

4.3 Formalizing interpretation

To represent interpretations we must formalize the complex package of structured information that figures in the speaker's communicative intention. Interpretations must specify the action taken, its description under the grammar, the preconditions that the action depends on and the outcome that the action is to achieve.

We formalize the action using a tree-adjoining grammar (TAG) (Joshi, Levy & Takahashi, 1975; Schabes, 1990), where any utterance is derived as the composition of elementary tree fragments that specify a lexical item and all of its syntactic arguments. For example, a schematic TAG derivation of *it's light brown* is as in (11).

(11)



We spell out the interface between structure and interpretation for structures such as (11) by formalizing the conceptual apparatus developed in Section 2.1. The grammar may associate any elementary tree with parametric actions that take their effects in the context whenever a speaker uses an utterance containing that elementary tree. The parameters in the action are subject to constraints that must be met in context. In addition, compositional semantics—here implemented via unification, following (Gardent & Kallmeyer, 2003)—may require trees to share parameters, so that one tree may function to specify the value of a free parameter in another.

The constraints associated with (11) are given in (12).

(12) $\texttt{predication}(M) \wedge \texttt{brown}(C) \wedge \texttt{light}(C)$
$\texttt{in-focus}(X)$

These constraints represent instructions to fill the three parameters, M, C and X, through which the update for this utterance will be constructed. For M we must find a relevant task action M that interlocutors normally achieve with a declarative sentence such as this; for C we must find a relevant color that can be described as *brown* and *light*; X is the object the utterance describes, and since the utterance uses *it* to refer to it, the grammar requires that *in-focus*(X) holds.

Note that these constraints link utterances to the contributions that are possible in a specific task. As in (DeVault, Rich & Sidner, 2004), a KNOWLEDGE INTERFACE mediates between domain-general meanings and the domain-specific ontology supported in a particular application. This allows us to build interpretations using domain-specific representations for referents, for task moves, and for the domain properties that characterize referents. In fact, our architecture allows utterances to trigger an open-ended range of domain-specific updates. We do not factor updates to the IS through an abstract taxonomy of speech acts—instead, utterances directly make domain moves, such as adding a constraint.

In (13), we schematize the actions effected by the utterance:

(13) $M(X, C(X))$
$\texttt{PutInFocus}(X)$

In other words, we know from the form of the utterance that it will make a contextually-relevant predication M that says that referent X has property C. Moreover, since referent X is realized as the subject, the utterance will ensure that X has (and retains) the pragmatic status of being in focus.

Together, (11), (12), and (13) constitute a description of the utterance *it's light brown* as an action: the description says what the action is, what preconditions it has, and what outcomes it achieves. These clauses thus constitute the core of a communicative intention.

To completely spell out a reason to use this action, however, we need to explain how exactly the preconditions of this action are met, and what outcome the action is intended to achieve. In this domain, we can link up (12) with specific instances in the information state of the conversation.

(14) $\texttt{predication}(\texttt{addcr}) \wedge$
$\texttt{brown}(\texttt{sandybrown}) \wedge$
$\texttt{light}(\texttt{sandybrown})$
$\texttt{in-focus}(t_1)$

These constraint instances simply say the following: first, that adding a constraint is the kind of action that in collaborative reference is achieved through a declarative statement; second, that in this context, this specific color, (RGB value F4-A4-60, or XHTML standard "sandy brown"), can be described as *light* and *brown*; and finally, that at this point in the conversation the target t_1 of the ongoing collaborative reference interaction is in focus. If the speaker commits to this specific link between the utterance and the context, the speaker must also commit to bringing about the specific changes to the domain state in (15).

(15) $\texttt{addcr}(t_1, \texttt{sandybrown}(t_1))$
$\texttt{PutInFocus}(t_1)$

In other words, the speaker must in fact add the constraint that the target of the current collaborative reference interaction is that specific color, and must keep that target in focus.

A system needs a complex structured representation—capturing the information in (11), (12), (13), (14), and (15)—to capture the content of a communicative intention. In (Stone & Thomason, 2002; Stone & Thomason, 2003), we show that abductive proofs do in fact constitute such complex representations.

To understand an utterance, we reason abductively to find a way of analyzing the utterance according to the grammar and recognize why (the speaker might think) the corresponding preconditions were satisfied. This gives us a proof that could represent a plan that could possibly be carried out. Before we use this plan to represent the speaker's intention, we must also check that the outcome the plan envisages is in fact a contextually-appropriate contribution to the ongoing task.

The possibility of tacit actions highlights the importance of our abductive characterization of interpretation. Consider understanding COREF's utterance *It's solid* as C_6 in

loop {

Perception

$e \leftarrow$ SENSE()

$\langle \hat{c}, i \rangle \leftarrow$ UNDERSTAND$(r, Z(c), e)$

$c \leftarrow$ UPDATE$(\hat{c}, i)$

Determination and Deliberation

$\hat{c} \leftarrow$ ACT-TACITLY(p, c)

$m \leftarrow$ SELECT$(p, \hat{c})$

$i \leftarrow$ GENERATE$(r, \hat{c}, m, Z(c))$

Action

ACT-PUBLICLY$(a(i))$

$c \leftarrow$ UPDATE$(\hat{c}, i)$

}

Figure 3: A general, computational specification for collaborating agents that take both tacit actions and public actions.

(9). In fact, COREF intends this utterance to have an analysis that is broadly analogous in form and content to the analysis of *It's light brown* that we've just seen. That is, the utterance gives more information about the domain object that serves as the value for target t_1 by indicating that the object is drawn with a fill rather than in outline. However, this interpretation is possible only if we ASSUME, abductively, that COREF has tacitly finished the clarification subdialogue that its last utterance contributed to. Otherwise, the appropriate action for the task would be to continue placing constraints to identify the color described as *light brown*.

As always, of course, we need to rule out this alternative interpretation as well as permitting the correct interpretation. COREF's implementation builds in some commonsense typechecking that does this. Such issues epitomize the detail and precision that is required in computational modeling, and underscore the need to characterize coordination in a substantive, computational way.

4.4 Modeling collaborative agency

We spell out agents' deliberation in pursuit of successful collaborative conversation in terms of this background of information states and communicative intentions by specifying agents' cognitive architecture and the computational mechanisms of coordination that they use.

The architecture we use is shown in Figure 3. It instantiates a general cognitive architecture for perception, deliberation and action (Russell & Norvig, 2003; Wooldridge, 2000) but specializes it for collaborative agency by introducing symmetric processes of understanding and generation and synchronized processes of update. Such architectures are motivated in greater detail in (Stone & Thomason, 2003) and (Stone, 2004b). The

specification presented in Figure 3 differs from earlier ones in allowing tacit actions, and maintaining the resulting uncertainties about context.

As always, it's best to explain the flow of control in this architecture by starting in the middle, with the deliberation that S does. When the deliberation begins, the state of the collaboration is represented as a checkpoint c. Now S has to decide how to proceed with the collaboration. To begin with, S determines and carries out whatever tacit actions are necessary at this stage of the collaboration, drawing on private information P.

The step:

$$\hat{c} \leftarrow \text{ACT-TACITLY}(p, c)$$

achieves this by updating the actual state of the collaboration from c to $\hat{c}$. Because S is the author these tacit actions, S knows that this update has taken place. However, these tacit actions introduce ambiguity for H and thus for the collaborators' mutual information. Before, the collaborators all mutually knew they were in state c; now all they mutually know is that they are in some state from $Z(c)$.

Next S has to choose an appropriate public action to perform based on private information P and the actual state of the collaboration $\hat{c}$. This choice is accomplished by the step:

$$m \leftarrow \text{SELECT}(p, \hat{c}).$$

This choice results in a schematic "message" m that S intends to convey; the exact way that this action will be carried out must be chosen to preserve coordination at the next checkpoint, and m will therefore leave details open that will have to be specified later.

The final step of deliberation is generation. The generator takes a set of resources r which models the available communicative conventions in the domain. It also reasons from the message m and the actual context $\hat{c}$. It will need to construct a way of carrying out m that makes sense in $\hat{c}$. But the generator also has to flesh out this action enough to enable the hearer not only to recognize m but $\hat{c}$ itself. That means the generator must take into account $Z(c)$, the horizon of alternatives that define the information state of the interaction. The generator now produces an intention i:

$$i \leftarrow \text{GENERATE}(r, \hat{c}, m, Z(c)).$$

This intention determines an action $a(i)$. Agent S now publicly carries out this action. Finally, to represent its effects, S updates the context from $\hat{c}$ using intention i:

$$c \leftarrow \text{UPDATE}(\hat{c}, i).$$

We now consider the beginning of the cycle of perception, deliberation and action from the standpoint of agent H. H is maintaining the initial checkpoint context as c, and so maintains the corresponding horizon $Z(c)$. H senses the action e that S has just made. H first understands e, using the communicative resources r and the contextual alternatives $Z(c)$:

$$\langle \hat{c}, i \rangle \leftarrow \text{UNDERSTAND}(r, Z(c), e)$$

The result has two parts. One is a representation of the context $\hat{c}$ in which S must have been acting, reflecting the tacit actions that S implicitly carried out. The other is a representation i of the intention that motivated S's action in the context. To reach the new checkpoint, H carries out the same update that S did:

$$c \leftarrow \text{UPDATE}(\hat{c}, i)$$

This context c is to be the next agreed checkpoint. It is now H's turn to act.

This framework allows us to be precise about what it is to maintain coordination in dialogue. Coordination maintenance is embodied in the principle:

> For all resources r, checkpoints c and states $\hat{c} \in Z(c)$, and messages m:
> If i is obtained by $\text{GENERATE}(r, \hat{c}, m, Z(c))$,
> then $\langle \hat{c}, i \rangle = \text{UNDERSTAND}(r, Z(c), a(i))$.
>
> **Principle of Coordination Maintenance**

The principle just says that, for any fixed perspective, any utterance $a(i)$ is understood as its speaker intends it. Suppose interlocutors share the background knowledge that fixes their communicative resources, their updates, and their horizons. Then if they share access to the initial conversational state, if they observe actions accurately, and if the coordination principle is satisfied, then they will continue to agree on each checkpoint.

This formal result points to the consistency of the kind of view we are advocating here. It shows that agents really could keep their view of context in alignment even without sharing all the information presupposed by their collaborators. Of course, the assumptions of the formal result are severe. Natural conversation succeeds even when there is more divergence in perspective and less certainty in action than our result considers. In this sense, the result calls attention to some of the idealizations and limitations in this provisional model. We leave more robust models for future treatments.

4.5 Coordination

The architecture of Figure 3 proposes substantive changes in how we think about and implement processes of understanding and generation to take tacit actions into account. Understanding no longer exploits a precisely specified context; there is uncertainty. Conversely, generation needs to make a single context salient in a range of alternatives. Moreover, as we have seen, these new processes must coordinate their activities.

Despite these changes for understanding and generation, we can carry over a substantial amount of conceptual and computational infrastructure into the new framework. To pursue the parallel, let us make explicit the interpretive mechanisms one normally thinks about, where all the relevant information is known. Such mechanisms apply, in effect, to an individual candidate state c from the horizon. They deliver the set of maximally plausible interpretations for an action e in that state. We will use the notation $I(c, e)$ for this set of interpretations. The calculation of $I(c, e)$ is just interpretation with complete information. To carry out this calculation, we might use the abductive approach to interpretation advocated in (Stone & Thomason, 2003) and the incremental use of grammatical

resources advocated in (Stone, 2004b). But there are certainly other ways to do it. (At the same time, we expect that an abductive approach to interpretation can also be extended to describe the understanding process in its entirety.)

As we hinted in Section 3, H must SIMULTANEOUSLY recognize which state from the horizon is actual and which interpretation action e has there. Here is a sketch of how this process might proceed. Agent H considers each state c in turn, computes $I(c, e)$ and inspects the result. It may be that $I(c, e)$ is empty. This means, for example, that action e would never be an appropriate continuation of the joint activity after the tacit actions that lead to c. Of course, this means that S could not have done those tacit actions; c can be eliminated. Alternatively, it may be that $I(c, e)$ has multiple elements. This means that e is ambiguous in c. This also means that c can be eliminated, because in c, S would not have fulfilled the obligations of coordination. If S had meant to convey one of these interpretations in c, he would have done e differently—for example, by adding words to disambiguate e. That leaves only states where $I(c, e)$ is unique. If there is exactly one such state c, then c must have been the actual state before the action was taken, and $I(c, e)$ must be the interpretation of the action there. Otherwise, since some sequences of tacit action are likely to be more obscure than others, it is convenient to assume a plausibility ordering over the states in the horizon. In this case, S must have meant the most plausible c for which $I(c, e)$ is unique. We'll use the notation $\text{BEST}(C)$ for the most plausible element of a set of states C.

We offer a formal summary of this discussion:

> Suppose $\langle \hat{c}, i \rangle = \text{UNDERSTAND}(r, Z(c), e)$
> Then $\hat{c} = \text{BEST}\{q \in Z(c) : |I(q, e)| = 1\}$
> and $\{i\} = I(\hat{c}, e)$.

The formalism makes clear that we can, in principle, carry out interpretation in an uncertain context with the computational ingredients we use for interpretation in fully-specified contexts.

Now let's consider how S plans an utterance. Suppose that S has a provisional utterance e in mind, and thereby intends to signal intention i with resources r in context $\hat{c}$ and horizon $Z(c)$. Suppose, however, that $\langle \hat{c}, i \rangle$ is not what results from $\text{UNDERSTAND}(r, Z(c), e)$. Then e is not a satisfactory utterance and has to be modified somehow. There might be either of two problems with e. Perhaps $I(\hat{c}, e)$ is not uniquely $\{i\}$. This is the familiar kind of ambiguity that can come up even without tacit actions. The generator needs to signal interpretive connections that link $\hat{c}$ to the intended interpretation i but that fail to link $\hat{c}$ to other alternative interpretations that are at least as salient as i. The generator can do this by adding words to e and corresponding presupposed content to i. Alternatively, perhaps there is a more plausible state c' for which $I(c', e)$ is unique. This ambiguity arises only because of tacit actions and the possibility of enlightened update. Nevertheless, the solution for the generator is basically the same as always. The generator has to add words to e and corresponding presupposed content that links $\hat{c}$ to the intended interpretation i but cannot be used to link c' to salient interpretations. In effect, just as we must distinguish i from its alternatives, we must distinguish $\hat{c}$ from its alternatives too.

Again, the formalism makes clear that we can, in principle, carry out generation in an

uncertain context with similar computational ingredients to those that fit fully-specified contexts. As always we can build a generator by planning—constructing utterances and their intentions incrementally, using steps of grammatical derivation that add syntax, semantics and pragmatics simultaneously (Stone et al., 2001). By consulting models of interpretation and understanding, the generator can ensure that it conforms to the principles of coordination required to align with others' understanding.

We can illustrate the kinds of reasoning we need in connection with dialogue (9). We have already seen that, in uttering *It's solid*, COREF calculates that the hearer will be able to recognize the tacit completion of the clarification subdialogue from the kind of assertion that COREF is making. But this is a complex prediction encapsulating a range of knowledge and assumptions about the hearer; it might turn out otherwise. If it did, COREF would choose to say something more explicit here, such as *The target square is solid*, which would impose additional grammatical preconditions and thereby unambiguously identify the target t_1 of the main collaborative reference task rather than the target S of the clarification subtask. In fact, we can see the effects of reasoning about coordination elsewhere in the utterances COREF does produce in (9). In generating C_2, for example, COREF iteratively elaborates its description from *brown* to *light brown* so as to distinguish the light brown of its intended target from the darker brown also visible on the display.

5 Elaboration and discussion

We can now revisit our knowledge-level account of presupposition in light of the specific implementation we have provided for it. Here we draw on the ideas we have developed to offer some perspective on matters of ongoing debate in the literature on presupposition accommodation. Recall that we view presuppositions as ingredients of a speaker's communicative intention and interpretation as interlocutors' coordinated recognition of these intentions. In Section 5.1, we contrast this view with other models of how presuppositions are recognized, and suggest that our model clarifies the place of accommodation in interpretive reasoning. At the same time, in Section 5.2, we argue that it also clarifies the kinds of data that require substantive explanation as exemplifying exceptional patterns of interpretation. We suggest that our computational approach is a necessary ingredient in explaining how interlocutors manage, exploit and learn to flout pragmatic rules as well as to follow them.

Of course, not all presuppositions involve planning, or depend on a well articulated context of conversational background and purposes. Our story remains provisional. We maintain only that if—as we believe—reasoning is a crucially important part of the story, then it is best to begin in a domain where reasoning is well behaved and can be formalized in detail. Then, if success is achieved in this domain, one attempts to extend the scope of the formalization. We suspect that, as in research on conversational implicature, too much attention has been paid to relatively exotic examples of presupposition accommodation, and that we might make better progress by concentrating on much larger bodies of everyday examples.

5.1 How accommodation takes place

David Lewis introduced a pretheoretic characterization of accommodation as an interpretive strategy for conversation.

(16) If at time t something is said that requires presupposition P to be acceptable [or: something is said that requires for its truth that P be presupposed], and if P is not presupposed just before t, then—*ceteris paribus* and within certain limits—presupposition P comes into existence at t.
(Lewis, 1979, p. 340)

A wide range of research has now fleshed out this characterization by clarifying and formalizing aspects of the reasoning involved.

For example, it is important that this accommodation can be a natural and transparent adjustment to the state of the conversation. In effect, accommodation makes sure that an utterance IS acceptable in the context in which it is interpreted. The process of understanding will show that P is required, and so will narrow down the alternatives to those that contain the information P. Thus when the hearer considers what response the utterance deserves, the utterance seems unproblematic. The fact that accommodation is in this sense an unreflective and inevitable side-effect of interpretation is emphasized by Stalnaker (Stalnaker, 1998) and von Fintel (von Fintel, 2000).

Moreover, we can directly abstract a formal model of accommodation from Lewis's characterization. There are possibilities and limits for what can be accommodated, and this suggests that we should formalize the state of the conversation in terms of a SET of alternatives. We formalize the inference that P is required as narrowing down this set of alternatives to those that contain the information P. This is the influential and attractive picture of Beaver (Beaver, 2001).

These insights and formal tools largely carry over to our approach. For example, when an agent characterized as in Figure 3 plans or understands an utterance, it consults its horizon, a set of alternatives each of which paints a different picture of the state of the conversation. An utterance can select one of these states by imposing grammatical requirements that fit it alone. This mirrors Beaver's formalization. The effect of this reasoning is to recognize the state of the conversation as a side-effect of the understanding process, much as Stalnaker and von Fintel describe it. In fact, as specified in Figure 3, the updates contributed by the utterance always advance to a state in which their preconditions are satisfied.

However, in our account, this reasoning is closely tied to an overall knowledge-level explanation of conversation as a collaborative interaction mediated by communicative intentions. The formal structures we use are independently-motivated representations of communicative intention and the reasoning processes we use solve independently-motivated problems of signaling and recognizing communicative intentions from utterances. Among other things, this means that when we apply our model to specify the behavior of a conversational agent or analyze examples of accommodation, we can work with a richer fabric of explanations. For instance, by thinking in terms of enlightened update, we can better understand how adjustments to the context as characterized by

Lewis in (16) can naturally arise for agents who coordinate successfully and collaborate effectively with one another.

Take Stalnaker's and von Fintel's description of the evolution of presupposition as a side-effect of understanding. One motivation for this account is to save a theory of presupposition that takes a speaker's presuppositions to be part of the common ground. Since presuppositions will be part of the common ground by the time the utterance is updated, the argument goes, it's RIGHT to view the grammar as imposing requirements of mutuality.

We do not see how to incorporate this idea of sequential update into an account that does justice to general pragmatic reasoning. Since the interpreter must find the best overall explanation of the utterance, presupposition resolution can't be an isolated subsystem but must interact with other sources of alternative interpretations. Consider, for instance, the interaction of presupposition with ambiguity resolution. Suppose there are TWO competing interpretations, each of which imputes a different state to the conversation by the grammatical requirements it imposes. Each interpretation is consistent and unique in its own imputed context. The hearer cannot use the accommodated information to decide between them. Instead, the hearer must fall back on the information that's ACTUALLY mutual—information from earlier in the conversation that creates a mutually-evident preference for one context and its corresponding interpretation. The speaker, in formulating utterances, must do likewise. This is why we continue to offer a knowledge-level characterization of the problem of interpretation as drawing on mutual supposition.

Thus, while you can narrate our system's reasoning in Stalnaker's and von Fintel's terms, you CANNOT thereby eliminate the fact that there are TWO fundamentally irreconcilable bodies of information at play. There is the information that's ACTUALLY shared, as represented in our system by the horizon as a whole. And there is the information in the state of the conversation through which GRAMMATICAL REQUIREMENTS can be satisfied, as represented in our system by a single state on the horizon. Of course, that state on the horizon WILL be identified in utterance understanding, so all the information it contains WILL be mutually supposed before utterance update. But to develop a theory of presupposition and accommodation that explains interlocutors' successful coordination, both utterance planning and utterance understanding need to track the actual current body of mutual suppositions, and not just optimistically anticipate the results of accommodation.

Similarly, consider Beaver's model of accommodation as a process of eliminating alternatives for the context. Beaver motivates a set of alternative contexts by imagining discourse produced by an uncooperative speaker who is ignorant of the audience—so that the speaker's interpretation task resembles that of an overhearer, for whom the discourse is not intended. In such a circumstance, while the grammar may force the speaker to act as though certain information is mutually supposed, the audience will be uncertain about what information the speaker is appealing to. They can find this out from the grammatical requirements his utterance imposes.

Again, at the knowledge level, Beaver motivates a different reasoning problem from ours. Our system has no uncertainty. It knows that what's mutual is ONLY what ALL states on the horizon have in common. It knows information that's present in some states on the horizon but not in others is NOT mutually supposed. If imposing a grammatical

requirement simply MEANS that information is mutually supposed, our system is WRONG to impose requirements that select some state on the horizon over others. It can't even be "flouting the rules", disobeying them to make a point; it doesn't have such a point. We characterize the problem our system solves differently. Our system, like any system that uses intentions to pursue collaborations, represents intentions in part through commitments about the preconditions of intended actions. It makes those commitments when they are consistent and coordinates them based on mutual supposition.

5.2 The problematic of accommodation

All this being said, we acknowledge that accommodation in general, including cases that are harder to characterize as enlightened update, is in many ways a puzzling and challenging phenomenon. In dealing with it at a broad theoretical or even a philosophical level, it is difficult to avoid being put in the position of postulating rules that are normally and routinely violated.

5.2.1 Accommodation on the cheap

Informative presupposition arises as a problem in the presence of a pragmatic rule requiring an utterance involving a presupposition to be appropriate only if its presuppositions are mutually supposed at that stage of the conversation. We are not committed to such a rule; the alternative rules that emerge from the ideas presented above would rather be (1) that an utterance involves a presupposition P if the intention underlying the utterance is committed to the presupposition, and (2) that an utterance is only appropriate to the extent that its presuppositions can be recognized and added to the common ground.

To take something on board in a conversation is not the same as to believe it. As Stalnaker points out (Stalnaker, 1975), the appropriate modality is "supposing for the sake of the conversation." A hearer can be credulous, coming away from a conversation fully believing it all, or can be skeptical and selective.

Often, we may be willing to take something on board in this way to avoid distracting the flow of the conversation. Consider the example of the ailing cat once again. You call me and say:

(17) I need to cancel our lunch meeting. I have to take my cat to the veterinarian.

On hearing this, I may be mildly surprised. In fact, I thought you were a dog lover. But the point of the conversation is our lunch meeting, and—even if I suspect you of making up an excuse—I would rather not confront you or call for a detailed explanation. I get on with the task of rescheduling the lunch, thereby taking the pet cat on board for the sake of the conversation.

Even if I'm not at all surprised or suspicious—you are, let's suppose, the sort of person I'd expect to own a cat—what I take on board at this point is (1) irrelevant to the main topic, and (2) underspecified. In effect, all I am adding are two almost naked discourse referents: a cat belonging to you and a veterinarian who is attending to the cat. They are not relevant to the topic, are not likely to be mentioned again, and I can forget them more or less immediately if I care to. The effort involved in this accommodation is minimal.

On the other hand, when information is topical and important, and certainly when it

is controversial, it is harder to slip it in by accommodation. If you call me asking me to pick up your cat at the veterinarian, you are unlikely to begin the conversation by saying

(18) Hello. Can you pick up my cat at the vet?

if I have no idea you have a cat. It is politer and more orderly to explicitly introduce the cat and its vicissitudes before making the request.

For accommodation to work well, it should be cognitively inexpensive.

5.2.2 The meaning of presupposition failure

But there are exceptions. For example, it is fairly common for the first sentence of a book to have presuppositions, and to violate normal rules about the orderly presentation of information. The first sentence of *The Missing World*,[5] for instance, is this:

(19) They were quarrelling on the phone when it happened, although anyone overhearing them might easily have failed to detect the fury that lay behind their pragmatic sentences.

Such cases do not always seem to be cognitively inexpensive. Nor do they seem to exemplify the kind of enlightened update we have explored, where speakers appeal to readily-identified real-world information to contribute naturally to discourse. Instead, examples of this kind are motivated and shaped by stylistic considerations, and intended to achieve an artistic effect. And they are genre specific; you would not expect a conversation or a tractor manual to begin with a sentence like this.

These considerations suggest that we must distinguish between enlightened update and other cases where a speaker uses an utterance whose presupposition P is not mutually known. One possibility (Clark, 1992) is that the speaker pretends to be somebody else, or to be addressing some other audience than is actually present, as in Clark's pragmatic reconstruction of storytelling, humor or irony in terms of "layering". In effect, then, the presupposition P characterizes the state of the pretend discourse, and the interpreter must recognize the pretense in order to understand the discourse. That seems to fit (19) from *The Missing World*.

Such explanations may require us to change the way we think of the problem of interpretation as a whole, not just the process of presupposition-accommodation. Now interpretation involves the open-ended reconstruction of the speaker's pretense, and allows the audience to explore a wide range of assumptions to make sense of the utterance. Even hypothetical alternatives to the preceding discourse can be considered. In fact, this is a common feature of jokes (Richie, 2003). And it may explain the pronouns in (19). Moreover, what we really take on board from a joke or story is only indirectly related to the context we reconstruct for it. Perhaps the assumptions give voice to to a real truth that we would not otherwise broach. Or perhaps the assumptions provide a foil to the way we think the world is. So the reasoning we need here is not just a matter of accommodating the speaker's presupposition. In many of these cases, only knowledge about the speaker can settle which, if any, was intended. In others, the intention may be incomplete, artificial, and obscure, and the problem of interpretation will become open-ended, conscious, and problematic.

[5] By Margot Livesey. Alfred A. Knopf, New York, 2000.

Karttunen's historically important example (a commencement announcement) is a similarly outlying case, in which the presupposition is the new information that the author intends to communicate.

(20) We regret that children cannot accompany their parents to commencement exercises.

Cases like these are more like appositive constructions than like ordinary presupposition sentences—that is, Sentence (20) is like

(21) Children, we regret, cannot accompany their parents to the commencement exercises.

Again, we suspect that this case is idiomatic. Perhaps these usages can be explained by a desire to save face. By using an utterance which (exceptionally) presupposes P to say that P, the speaker avoids the need to overtly to raise the question whether P as a topic of discussion (Thomason, 1990). Moreover, the kinds of presuppositions that can be made mutually available through this strategy, while perhaps broader than those that result from tacit actions, remain quite limited; for example, one cannot retrospectively change what utterances have been made in the conversation (Beaver & Zeevat, 2004). Corpus evidence would be one way to confirm our guess that usages like these are stylistically marked, and are confined to a few "presupposition triggers."

5.2.3 Two knowledge levels

How then should we think of exceptions? Perhaps we need to distinguish between modeling the reasoning itself and giving a rational reconstruction of the principles that make the reasoning coherent.

Allan Newell's "The knowledge level" was the first AAAI Presidential address, delivered at Stanford University in 1980.

Newell took as his starting point the emergence within computer science of a number of levels of from which it could be fruitful to describe a computational system, with each level characterized by its own ontology and laws. He lists five levels, including the CIRCUIT LEVEL and the SYMBOLIC or PROGRAM LEVEL. While a computer engineer might use the former level (ignoring the program entirely and concentrating on the flow of digital information through the registers of the device) to analyze the behavior of a computer running a simple program, a software engineer, taking the correct behavior of the hardware for granted, might concentrate on the transitions between symbolic states that are induced by the program.

To these levels, Newell suggested the addition of a higher level, involving the attribution of BELIEFS and GOALS to the system. This idea is congenial to computer scientists, since it is natural and unavoidable to think even of simple of systems as having goals and beliefs. For instance, it is natural to explain why my navigation system is beeping by saying it thinks I have reached a turn in my route and wants to alert me to turn. But the idea has also been adopted by some philosophers; see, for instance, (Dennett, 1979).

In his later work, especially in (Newell, 1992), Newell elaborated these ideas and argued for their general applicability in cognitive science. There, he also speculates about

other levels, including a social level. He doesn't provide details about the sort of laws that would be appropriate for the social level, but presumably (like the knowledge level, which could make use of a variety of formalization devices, including probability functions and more qualitative logical formalizations) the social level could be formalized using ideas from macroeconomics, or using a logic providing for the formalization of group norms.

This distinction can be useful in thinking about accommodation. Anyone who takes accommodation or similar phenomena seriously is likely to be plagued by an apparent appeal to a rule that is then undermined by the accommodating practice. (And it is hard not to take it seriously as a general phenomenon, on seeing the range of examples that Lewis lays out in (Lewis, 1979).) The undermining, which affects Lewis' account in (Lewis, 1969), as well as Grice's idea, formulated in (Grice, 1989, pp.30ff) of how implicatures are induced by "flouting" a rule of conversation, conforms to the following generic pattern.

(22) A general rule R of the form *Do A only if C* is postulated, where C is a condition that can be manipulated by an audience H. An agent S is observed by H to be doing A while C is mutually known to be false. H then acts to make C true, and s expects H to so act.

To take a nonlinguistic example, a woman at a restaurant violates the rule (R1) "Don't sit down unless there is an empty chair positioned behind you," knowing that a waiter will position the required chair, and that indeed her sitting will be the signal for this act. Here, it seems natural to say either (taking a nonmonotonic view of the rule) that R1 has not been violated, because it is defeasible and this is one of the exceptional conditions, or (taking a monotonic view) that R1 has to be replaced with an improved, qualified version: "Don't sit down unless there will be an empty chair positioned behind you when it is needed."

Von Fintel uses the monotonic tactic in (von Fintel, 2000) to defend Stalnaker's rule (R2) "Do not trigger a presupposition P unless P is in the common ground that is available for updating the conversation." The idea is that, like the waiter, the audience can add the required presupposition to the common ground before updating with the content of the utterance.

Von Fintel seems to treat pragmatic rules like Stalnaker's as indefeasible—that is, he replaces R2 with a more accurate, qualified rule. But a defeasible approach to pragmatic rules might be more explanatory. In the restaurant scenario, it is natural to say that the woman is acting as if there were a chair behind her. But it is hard to explain how she could be doing this unless R1 has some force as a rule of conduct. Also, giving R1 some force as a *ceteris paribus* rule can also provide a more plausible reconstruction of the woman's and the waiter's reasoning. If someone were to feel faint in a restaurant, and involuntarily begin to collapse into a sitting position, a waiter might well rush to the spot with a chair. But in the accommodation scenario, we have a strong intuition that the woman MEANS SOMETHING by sitting, and that the waiter has correctly recognized this meaning. If we want to provide a rational reconstruction of this reasoning (rather than treating it as a special convention that applies in some restaurants), it will be useful to give rulelike force to R1.

Similarly, taking R2 to be defeasible could help to explain Stalnaker's claim that when s says "I have to pick up my cat at the veterinarian," s is acting as if it were part of the common ground that s has a cat.[6]

But the analogy between accommodation in conversation and in the restaurant, and attempts like von Fintel's to neutralize the undermining effect of accommodation, are imperfect in cases in which accommodation leads to routine violations of the rule. As many people have pointed out, violations of R2 occur with great frequency, and there is no intuition in such cases that any rule or expectation has been violated. This makes it difficult to give R2 the status of a defeasible rule—at least, if we expect defeasible rules to hold usually, or normally. The rule we need to invoke to explain the accommodation really seems to be undermined when accommodation becomes the rule.

Here it may be useful to think of R2 not as a knowledge-level rule governing behavior, but as a rule at the social level. It is possible for a rule to have acknowledged status as a social norm of a fairly idealized sort, and yet for it to be routinely flouted without any sense of violation. "Don't interrupt" might be an example, in contemporary American society. If we assume a society of accommodating agents, who are also capable of learning and forming habits, we would expect the rules that give rise to instances of accommodation, over time, to give rise to habits that flout these rules. But this doesn't mean that these rules will not preserve a vestigial status, or that they won't be useful for certain theoretical purposes.

Social rules can certainly be violated in order to achieve an effect—think of refusing to shake someone's hand when it is offered—and when the effect is cooperative and expected, and matches a familiar and habitual pattern, it would be natural to expect that there should be no perception of abnormality, or that a rule has been violated.

It may seem quixotic to invoke rules that are routinely and normally violated, but idealized social rules may have a useful role to play in rational reconstructions, like Grice's rationale for implicatures and a reconstruction of the sort that Stalnaker (Stalnaker, 1998) offers of Lewis' account of presupposition accommodation. And, it seems we people can have intuitions about such vestigial rules that are quite robust. Quite possibly, vestigial rules of this sort could play a useful expanatory role in a systematic theory of pragmatic norms—providing coherence, motivating or even enabling the derivation of lower-level rules, and explaining how these rules are learned. However, since we lack a systematic theory of this kind, these suggestions are only tentative.

5.2.4 Explanation and implementation

In any case, it would be a mistake to invoke rules of this sort in a knowledge-level specification of a conversational algorithm. A conversational planner s has the goal, say, of updating the common ground with an individual c, which is a cat and belongs to s, with an individual v who is a veterinarian, and with the proposition that s has to pick up c at v's office. There are various ways of accomplishing this goal, including:

[6] In (Stalnaker, 1975, p. 202), Stalnaker suggests that a speaker is pretending that the presupposition is part of the common ground in such cases. Although there are cases, like (19), that are best described as pretenses, Stalnaker's cat example (3) is not one of these. Here, it is probably better to say that the speaker is acting as if it is part of the common ground.

(23) I have a cat.
The cat is at a veterinarian.
The veterinarian has an office.
I have to pick up the cat at the veterinarian's office.

(24) I have a cat.
I have to pick up the cat at the veterinarian's.

(25) I have to pick up my cat at the veterinarian's.

The planner needs to know that such alternatives will accomplish the same result, and to know that normally the last version is the preferred, unmarked way of doing this. But to reconstruct the planner's knowledge and reasoning, we do not need to postulate a rule prohibiting a definite reference to an individual not already in the common ground. Indeed, such a rule could quite likely get in the way of using the unmarked formulation. A rational reconstruction is not the same thing as a knowledge-level specification.

In the recent literature on presuppositions, it has been suggested that many presuppositions arise as conversational implicatures; see, for instance, (Simons, 2001). We do not need to take a stand on this issue here, as long as Simon's suggestion is thought of as a high-level explanation of presupposition, perhaps at the social level. At the implementation level, and even perhaps at the knowledge level, it would be difficult or impossible to account for presuppositions by reasoning processes that appealed to first principles rather than specific grammatical conventions and specific contextual resources. Implicatures can in principle be derived from first principles using abductive reasoning, but the knowledge that would be needed to support the processes is very open-ended. A general solution to the problem of conversational implicature would probably require robust, general-purpose common-sense reasoning. In any case, even if many or even all presuppositions originate as conversational implicatures, many of them would be highly *generalized* implicatures, ones that hold over a large set of interpretation contexts. In many reasoning domains, it would be a reasonable idealization to treat these implicatures as having a grammatical origin.

6 Looking ahead

Pragmatics involves both nonlinguistic and linguistic (grammatical) information and reasoning. But there is a sort of exclusion principle that applies to research projects in pragmatics—the projects that do justice to grammar tend to neglect common-sense reasoning, and those that do justice to nonlinguistic reasoning tend to neglect grammar. Our paper may appear to be of the second kind: not only have we said nothing about presupposition triggers, presupposition projection, and the compositional calculation of presupposition from grammatical knowledge, but we have said very little about the typical examples that are considered in the presupposition literature.

Our focus on examples involving tacit actions was motivated by the need to stay close to areas that we know how to formalize and implement. But we believe that the ideas can be plausibly extended to more familiar cases. As for grammar, the framework we work

with is compatible with familiar grammar formalisms, and our implementations are readily integrated with a grammar component that specifies syntactic structures. We explore such connections elsewhere, particularly in (Stone & Thomason, 2002; Stone, 2004b).

Although we believe that linguistic theories of presupposition also could be combined with the reasoning formalism we have advocated, it remains to be seen how easy or useful it would be to carry this out on a large scale. We have yet to engage the challenges that occupy applied research in natural language processing, such as the practical acquisition of massive amounts of domain knowledge and rich, detailed grammars; the automated induction of preferences that predict interpretations accurately; and the engineering of implementations that can work with detailed grammars, vast knowledge bases and detailed interpretive preferences with reasonable computational demands on computational space and time. These issues would need to be addressed in any large-scale system, whether or not it incorporates presuppositions.

Returning to the cooking example, suppose that Bess has the initiative. Unseen by Andy, she fills a pot with water, puts it on a burner, and turns the heat on. She then says to Andy:

(26) When the water boils, we should blanch the tomatoes.

Her utterance is a proposal for the group to undertake a certain course of action. Such speech acts can be incorporated in the model by including constraints on future courses of action in the state of the system; these commitments can be formalized as sets of action sequences. Also, the model would need to be generalized to provide for exogenous changes of the domain state, such as a pot coming to a boil.

To successfully interpret this sentence, Andy must "accommodate" both the reference of *the water* and the presupposition that the water will boil. This means that he should understand the utterance as (1) a signal to augment his inventory of objects to include a quantity of water, (2) a signal to revise his view of the mutual activity to include a (tacit) action of Bess's putting this water to boil on the stove, and (3) a proposal to commit to blanching the tomatoes in this water when it boils.

Assume for the moment that the grammar associates Sentence 26 with a presupposition to the effect that W will boil, where W is a to-be-resolved quantity of water. This provides a natural and appropriate input for the interpretive task, which requires a preference for action sequences including a tacit action that provides a precondition for W's boiling. Grammatical presuppositions will guide the reasoning task appropriately in this case if the presuppositions correlate with preconditions that have been supplied by tacit actions.

In contrast, a discourse like

(27) We need to blanch the tomatoes. We should boil some water.

proposes a course of action in which a quantity of water W (to be created by an as-yet unperformed action) is put in a container, heated, and used to blanch the tomatoes when it boils.

Recall that abductive reasoning requires nonpropositional information, in the form of preferences for certain explanations. The reasoning context of an inductive interpreter

will consist not just of a set of worlds, but of a space of alternative explanations, together with preferences over these explanations. In (Stone & Thomason, 2003) we developed the idea that utterances have conventional dynamic effects on this context, and used this idea to model certain attentional effects. Roughly, for instance, an indefinite noun phrase is associated with interpretations that create a new reference, rather than searching for a reference that has already been created.

Regarding the role of presupposition in collaborative planning domains, we want to suggest that presuppositions correlate with tacit actions to which the intention behind the utterance is committed. They do this by adjusting preferences for interpretations so that explanations that associate the content of the presupposition with tacit actions are less costly. For instance, the declarative content of *When the water boils, you will blanch the tomatoes* is that H will blanch T when W boils; its interpretive force is a preference for intentions that commit S to having performed actions that create the preconditions for W boiling.

We feel that this suggestion has a certain plausibility, but it needs to be tested by incorporating it in a dialogue system and evaluating the system's implementability and performance. The fact that such ideas can be tested in this way is, we believe, an important contribution of A.I. methodology to pragmatics, an area which desperately needs more effective ways of confirming and disconfirming theories.

References

Abusch, D. (2005). Triggering from alternative sets and projection of pragmatic presuppositions. Unpublished manuscript, Cornell University. Available at http://semanticsarchive.net/Archive/jJkYjM3O/Abusch-Triggering.pdf.

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics*, *4*(6), 1236–1239.

Barker, C. (2002). The dynamics of vagueness. *Linguistics and Philosophy*, *25*(1), 1–36.

Beaver, D. (2001). *Presupposition and Assertion in Dynamic Semantics*. Stanford, California: CSLI Publications.

Beaver, D. I. & Zeevat, H. (2004). Accommodation. Available at http://montague.stanford.edu/ dib/Publications/accommodation.pdf. To appear in G. Ramchand and C. Reiss (eds.), *Oxford Handbook of Linguistic Interfaces*, Oxford University Press.

Bratman, M. E. (1987). *Intentions, Plans and Practical Reason*. Harvard University Press.

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(6), 1482–1493.

Burge, T. (1973). Reference and proper names. *The Journal of Philosophy*, *70*(14), 425–439.

Clark, H. (1992). *Arenas of Language Use*. Chicago: University of Chicago Press.

Clark, H. H. & Marshall, C. R. (1981). Definite reference and mutual knowledge. In Joshi, A., Webber, B., & Sag, I. (Eds.), *Elements of Discourse Understanding*, pages 10–63. Cambridge, England: Cambridge University Press.

Herbert, H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.

Dale, R. & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233–263.

Dennett, D. C. (1979). Artificial intelligence as philosophy and as psychology. In Ringle, M. (Ed.), *Philosophical Perspectives in Artificial Intelligence*, pages 57–80. Atlantic Highlands, New Jersey: Humanities Press.

DeVault, D., Kariaeva, N., Kothari, A., Oved, I., & Stone, M. (2005). An information-state approach to collaborative reference. In Nagata, M. & Pedersen, T. (Eds.), *ACL 2005 Proceedings Companion Volume. Interactive Poster and Demo Track*. Association for Computational Linguistics. URL: http://www.cs.rutgers.edu/ mdstone/pubs/acl05ip.pdf.

DeVault, D., Rich, C., & Sidner, C. L. (2004). Natural language generation and discourse context: Computing distractor sets from the focus stack. In Barr, V. & Markov, Z. (Eds.), *17th International FLAIRS Conference (FLAIRS-2004)*, Menlo Park, California. AAAI Press.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about Knowledge*. Cambridge, Massachusetts: The MIT Press.

Gardent, C. & Kallmeyer, L. (2003). Semantic construction in feature-based TAG. In *Proceedings of the 10th Meeting of the European Chapter of the Association for Computational Linguistics*.

Geanakopolos, J. (1994). Common knowledge. In Aumann, R. & Hart, S. (Eds.), *Handbook of Game Theory, with Economic Applications, Vol. 2*, chapter 40. Amsterdam: North-Holland.

Grice, H. P. (1957). Meaning. *Philosophical Review*, *66*, 377–388. Republished in H.P. Grice, *Studies in the Way of Words*, Harvard University Press, 1989.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, Massachusetts: Harvard University Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274–307.

Heeman, P. & Hirst, G. (1995). Collaborating on referring expressions. *Computational Linguistics*, *21*(3), 351–382.

Hobbs, J., Stickel, M., Appelt, D., & Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, *63*(1–2), 69–142.

Joshi, A. (1982). Mutual beliefs in question-answer systems. In Smith, N. (Ed.), *Mutual Knowledge*, pages 181–197. London: Academic Press.

Joshi, A. K., Levy, L., & Takahashi, M. (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences*, *10*, 136–163.

Kaplan, D. (1989). Demonstratives: an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In Almog, J., Perry, J., & Wettstein, H. (Eds.), *Themes from Kaplan*, pages 481–563. Oxford: Oxford University Press.

Kripke, S. (1972). Naming and necessity. In Harman, G. & Davidson, D. (Eds.), *Semantics of Natural Language*, pages 253–355. Dordrecht: D. Reidel Publishing Co.

Larsson, S. & Traum, D. (2000). Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering*, *6*, 323–340.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge, Massachusetts: Harvard University Press.

Lewis, D. K. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(3), 339–359.

Lochbaum, K. E. (1998). A collaborative planning model of intentional structure. *Computational Linguistics*, *24*(4), 525–572.

Mackworth, A. (1987). Constraint satisfaction. In Shapiro, S. (Ed.), *Encyclopedia of Artificial Intelligence*, pages 205–211. John Wiley and Sons.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman.

Newell, A. (1981). The knowledge level. *Artificial Intelligence Magazine*, *2*(2), 1–20.

Newell, A. (1992). *Unified Theories of Cognition*. Cambridge, Massachusetts: Harvard University Press.

Perrault, C. R. & Allen, J. F. (1980). A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, *6*(3–4), 167—182.

Poesio, M. & Traum, D. R. (1997). Conversational actions and discourse situations. *Computational Intelligence*, *13*(3), 309–347.

Pollack, M. (1990). Plans as complex mental attitudes. In Cohen, P. R., Morgan, J., & Pollack, M. (Eds.), *Intentions in Communication*, pages 77–103. Cambridge, Massachusetts: MIT Press.

Pollack, M. (1992). The uses of plans. *Artificial Intelligence*, *57*(1), 43–68.

Purver, M. (2004). *The Theory and Use of Clarification Requests in Dialogue*. Ph.d. dissertation, King's College, University of London, London. Available at http://www.stanford.edu/~mpurver/papers/purver04thesis.pdf.

Putnam, H. (1975). *Mind, Language and Reality: Philosophical Papers, vol. 2*. Cambridge, England: Cambridge University Press.

Reiter, R. (2001). *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. Cambridge, Massachusetts: The MIT Press.

Rich, C., Sidner, C. L., & Lesh, N. (2001). COLLAGEN: Applying collaborative discourse theory to human-computer interaction. *Artificial Intelligence Magazine*, *22*(4), 15–26.

Richie, G. (2003). *The Linguistic Analysis of Jokes*. London: Routledge.

Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguistics and Philosophy*, *26*(3), 287–350.

Russell, S. & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach* (2 ed.). Englewood Cliffs, New Jersey: Prentice Hall.

Schabes, Y. (1990). *Mathematical and Computational Aspects of Lexicalized Grammars*. PhD thesis, Computer Science Department, University of Pennsylvania.

Simons, M. (2001). On the conversational basis of some presuppositions. In Hastings, R., Jackson, B., & Zvolenszky, Z. (Eds.), *Proceedings from Semantics and Linguistic Theory XI*, pages 431–448, Ithaca, New York. Cornell University.

Simons, M. (2003). Presupposition and accommodation: Understanding the Stalnakerian picture. *Philosophical Studies*, *112*(3), 251–278.

Sperber, D. & Wilson, D. (1995). *Relevance, Second Edition*. Oxford: Blackwell.

Stalnaker, R. (1998). On the representation of context. *Journal of Logic, Language, and Information*, *7*(1), 3–19.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*(5–6), 701–721.

Stalnaker, R. C. (1973). Presuppositions. *Journal of Philosophical Logic*, *2*(4), 447–457.

Stalnaker, R. C. (1975). Pragmatic presuppositions. In Munitz, M. K. & Unger, P. (Eds.), *Semantics and Philosophy*, pages 197–213. New York: Academic Press.

Stalnaker, R. C. (1981). Assertion. In Cole, P. (Ed.), *Radical Pragmatics*. New York: Academic Press.

Stickel, M. E. (1991). A Prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. *Annals of Mathematical Artificial Intelligence*, *4*, 89–105.

Stone, M. (2004a). Communicative intentions and conversational processes in human-human and human-computer dialogue. In Trueswell, J. & Tanenhaus, M. (Eds.), *World Situated Language Use: Psycholinguistic, Linguistic, and Computational Perspectives on Bridging the Product and Action Traditions*, pages 39–70. Cambridge, Massachusetts: The MIT Press.

Stone, M. (2004b). Intention, interpretation and the computational structure of language. *Cognitive Science*, *5*.

Stone, M., Doran, C., Webber, B., Bleam, T., & Palmer, M. (2003). Microplanning with communicative intentions: The SPUD system. *Computational Intelligence*, *19*(4), 311–381.

Stone, M. & Thomason, R. (2003). Coordinating understanding and generation in an abductive approach to interpretation. In Kruijff-Korbayová, I. & Kosny, C. (Eds.), *Diabruck 2003: Proceedings of the 7th Workshop on the Semantics and Pragmatics of Dialogue*, pages 131–138. Saarbrücken: Universität des Saarlandes.

Stone, M. & Thomason, R. H. (2002). Context in abductive interpretation. In Bos, J., Foster, M. E., & Mathesin, C. (Eds.), *EDILOG 2002: Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialogue*, pages 169–176. Edinburgh: Cognitive Science Centre, University of Edinburgh.

Thomason, R. (1990). Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In Cohen, P. R., Morgan, J., & Pollack, M. (Eds.), *Intentions in Communication*, pages 326–363. Cambridge, Massachusetts: MIT Press.

Thomason, R. H. (2000). Modeling the beliefs of other agents. In Minker, J. (Ed.), *Logic-Based Artificial Intelligence*, pages 375–473. Dordrecht: Kluwer Academic Publishers.

Traum, D. R. & Allen, J. F. (1994). Discourse obligations in dialogue processing. In Pustejovsky, J. (Ed.), *Proceedings of the Thirty-Second Meeting of the Association for Computational Linguistics*, pages 1–8, San Francisco. Association for Computational Linguistics, Morgan Kaufmann.

von Fintel, K. (2000). What is presupposition accommodation? Unpublished manuscript, Linguistics Department, MIT.

Wooldridge, M. J. (2000). *Reasoning about Rational Agents*. Cambridge, England: Cambridge University Press.